Three Essays in Applied Microeconomics

Omid Bagheri

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Thorwald N. Tideman, Co-chair
Klaus Moeltner, Co-chair
Kwok Ping Tsang
Wen You

August 10, 2017
Blacksburg, Virginia

Keywords: Immigration, Wage Differentials, STEM, Air Pollution, Respiratory Health

Three Essays in Applied Microeconomics

Omid Bagheri

ABSTRACT

This dissertation includes three chapters. The first and second chapters are related to economics of immigration, and the last one is about environmental economics. The first chapter studies people who work and live in the US on work visas such as H-1B, and compares them to natives. In this chapter I examine whether or not there exists any wage premium for or against college graduate immigrants who are on work visa compared to college graduate natives. I also check for any change of such a premium from 2003 to 2010. On the contrary to the common belief that foreign workers are cheap labor force, my results show that skilled immigrants holding temporary work visas on average have a significant wage premium over natives, and this premium has even increased significantly from 2003 to 2010 (from 14% to 22%). My results show that such a wage premium is different for men, women, and countries of origin, but I find no evidence supporting different premiums across different fields of study.

The second chapter of this dissertation studies the dynamics of the earnings gap between those immigrants and US-born individuals with bachelor's degrees or higher in science and engineering fields. The research question is that in case a gap exists for or against immigrants, how is it changing with the amount of time immigrants reside in the United States? I employ cross-sectional and longitudinal approaches to answer this question, and study the earnings gap between three groups of immigrants (based on the current residency status) and natives at entry and over time. I also compare natives with immigrants who migrated to the United States on different types of visas (permanent residence visa, work visa, study visa, and dependent visa). Results show that, upon arrival, immigrants on average have a considerable premium over the US-born, and this gap, surprisingly, even gets bigger with an approximate rate of 0.25% for the first 5-10 years of immigrants' residence in the US. This phenomenon could be due to the higher level of abilities and motivation among immigrants compared to natives. Another reason can be the selectivity among immigrants, meaning that more successful stays and others return. Unfortunately, due to the lack of information in data regarding these issues, they could not be controlled for in my models.

The last chapter is about environmental economics. This chapter exploits a daily time series data on pollen count and $PM_{2.5}$ level from 2009 to 2015 to study the separate impacts of $PM_{2.5}$ and pollen on the number of total, in-patient, and out-patient respiratory hospital admissions within different age groups in the Reno/Sparks metropolitan area in Northern Nevada. The results show that although $PM_{2.5}$ has a positive impact on the number of out-patient admissions in most of the age groups, there is no evidence that shows any relationship between the pollen count and the number of in-patient or out-patient respiratory admissions.

Three Essays in Applied Microeconomics

Omid Bagheri

GENERAL AUDIENCE ABSTRACT

This dissertation includes three chapters. The first and second chapters are related to immigration and its economic consequences, and the last one is about an environmental issue. My first chapter studies educated people who work and live in the US on work visas such as H-1B and compares them to educated US-born individuals. Any change of such a wage difference from 2003 to 2010 is also studied. On the contrary to the common belief that foreign workers are a cheap labor force, I find that skilled educated immigrants holding temporary work visas on average have higher salaries compared to natives and this wage gap in favor of immigrants has even increased from 2003 to 2010 (from 14% to 22%). This wage difference between natives and immigrants is different among men and women and also by immigrants' country of origin. However, I find no evidence which shows wage differences across different fields of study.

The second chapter is about finding the wage difference between US-born individuals and immigrants with college degrees in science and engineering fields of study at the time of their entry to the US, and more importantly studying the changes of this wage gap with more time immigrants reside in the United States. For this purpose, immigrants are grouped based on their current residency status. I also compare natives with immigrants who migrated to the United States on different types of visas (permanent residence visa, work visa, study visa, and dependent visa). Results show that, upon arrival, immigrants make more than the US-born, and this gap even gets bigger with an approximate rate of 0.25% for the first 5-10 years of immigrants' residence in the US. This could be because of the higher level of abilities and motivation among immigrants compared to natives. Another reason can be that more successful immigrants stay and others go back home or migrate to another country. Unfortunately, my data does not provide me with information regarding either one of these issues.

The last chapter is about environmental economics. In this chapter I use daily data on pollen (a type of allergen released in the air by plants) and $PM_{2.5}$ (an air pollution factor mainly caused by wildfires) from 2009 to 2015 and study the impacts of $PM_{2.5}$ level and pollen count on the number of respiratory related hospital admissions in the Reno/Sparks metropolitan area in Northern Nevada. I find that an increase in $PM_{2.5}$ leads to a bigger number of out-patient hospital admissions in most of the age groups. However, no evidence was found which shows any relationship between the pollen count and the number of in-patient or out-patient hospital admissions.

Acknowledgments

This work would not have been completed without the help and support of the many people who believed in me, first and foremost my co-advisors, Drs. Klaus Moeltner and Nicolaus Tideman. They never gave up on me and supported me when I needed it most. Thank you very much for all that you did for me and all that you taught me. I will forever be in your debt. I would like to express my most sincere acknowledgments to the other members of my dissertation committee, Drs. Kwok Ping (Byron) Tsang and Wen You, for their generous help and valuable support. I would also like to offer special thanks to Drs. Richard Cothren and Steve Trost for being such caring and supportive faculty members.

My most sincere appreciation goes to my parents Vida and Hossein, to whom I owe whatever I am today and whatever I have done. I also thank my brother Navid greatly for his bottomless support and endless love. I dedicate this work to the three of them.

At last, I would like to thank my very good friend, Mariya Tsyglakova, who was beside me whenever I felt low and needed support.

Contents

Chapter 1

Are College Graduate Immigrants On Work Visa Cheaper Than Natives?

1.1 Introduction

Some people believe that foreign skilled workers are a cheap source of labor for the United States. In other words, there is a general belief that there exists some wage discrimination against immigrants who are working with working permits (H-1B visa or other working visas) compared to the natives in the United States labor market.

It is certain that because of some sponsorship costs and some federal regulations entering the job market and getting hired is on average harder and more complicated for people with any visa compared to permanent residents (Green Card holders) or citizens. Employers obviously prefer to hire people who cost them less in terms of sponsorship or any other legal expenditure while qualified enough for the job. They also prefer natives or at least permanent residents since by hiring them (instead of foreign high-skilled workers with similar qualifications) they could count on longer cooperation with less legal difficulties. Yet, when a non-permanent resident individual gets a job and acquires a working visa after lots of competition and effort, the case will be different. Now, unless some econometric analysis is done, one cannot say for sure whether or not any significant wage difference exists between natives and those on work visas.

One aspect of globalization relates to outward mobility of work to foreign workers at remote locations as reflected in outsourcing or off-shoring of business processes and services, while another aspect of globalization relates to the inward mobility of foreign workers who are immigrants or on a work visa (Mithas and Lucas Jr, 2010). Currently, there are some different temporary visa programs in the US. One of the largest programs amongst all is the H-1B program that allows US businesses to temporarily hire high-skilled foreign workers (with at least a bachelor's degree) whenever there is a shortage of skilled professionals in specific professions especially in the Science, Technology, Engineering, and Mathematics (STEM) fields in order to improve economic growth and design innovative goods and services.

Recently, there have been many public discussions about immigration of the skilled workers to the United States and whether H-1B visa recipients are really highly skilled or not. A significant part of such discussions is about the advantages or disadvantages of the United States' H-1B visa program for the US labor market and the optimum annual cap on the number of H-1B's issued per year (which is set by the US Congress). On one hand, US businesses and the whole economy could benefit from hiring skilled foreigners on work visas. On the other hand, new high-skill immigration might have a negative effect on the wages of both foreign and native professional workers currently in the labor market. The H-1B program has been criticized for having a negative effect on native workers, interfering with the efficiency of labor markets by reducing the mobility of the workers on such visas, and providing employers an opportunity to exploit immigrant workers (Lofstrom and Hayes, 2011). The big concern is whether H-1B's are less paid compared to native workers or not.

Setting efficient visa caps is highly important because setting them too low may constrain domestic industries and usher in moving such industries to locations out of the United States. For instance, Microsoft established a complex in Canada for 150 foreign professionals that it could not bring to the United States due to restrictive H-1B visa caps in 2007. However, setting the annual caps too high can lead to some concerns about protecting American jobs and wages (Mithas and Lucas Jr, 2010).

Some other temporary work visa programs for high-skilled workers are L1 (for intra-company transferees), O1 (for individuals with extraordinary abilities), and TN (for skilled workers from Other NAFTA (North American Free Trade Agreement) countries. These programs are generally less controversial than the H-1B program (Lofstrom and Hayes, 2011). This is probably for two reasons: 1- the number of H-1B's is bigger and consequently this could have a more significant impact on the US labor market, and 2- H-1B could more easily be changed to a green card.

The 2003 and 2010 waves of National Survey of College Graduates (NSCG) are used in this study to examine how well immigrants working in the United States with work visa and at least a bachelor's degree are getting paid compared to college graduate natives. In this study, I check for any significant wage disparities between the two groups controlling for socio-economic and demographic characteristics of individuals. Moreover, I study the change of such a wage premium (for or against immigrants holding work visas compared with natives) from 2003 to 2010. I also analyze to see whether this wage premium is different for men and women, for individuals originally from some specific countries which send or has sent a significant number of immigrants to the United States, or for people having different fields of study. This study is the first of its kind using NSCG data to calculate the above mentioned premium in general and for different types of individuals, and also to observe the trend of changes in the premium over time.

The National Survey of College Graduates (NSCG) is developed by National Science Foundation (NSF) in collaboration with the Census Bureau of the United States. This dataset contains individuals who at least hold a bachelor's degree in any field from a university inside or outside the United States. The advantage of this dataset is that it has all the socio-economic information about immigrants and natives that are needed to study and compare the wages including the citizenship and detailed residency status of each person. Unfortunately, the dataset does not descry the exact work visa that an individual has. However, since all of the individuals in the sample are college graduates, and H-1B has a big share in the total number of work visas issued each year in the US, one will not be far off if assumes a big portion of the work visas are H-1B's.

After controlling for most of the factors that are affecting earnings, I find that contrary to the common belief that foreign workers are cheap labor force, after controlling for socio-economic and demographic characteristics, skilled immigrants holding temporary work visas have a 22 log point wage premium over American natives. The results also show that the premium has even significantly increased from 2003 to 2010 (from 14% to 22%). According to results, this wage premium is less for men compared to women and different for individuals originally from some specific countries like Canada, China, and Iran, but there is no evidence supporting different premiums for people having different fields of study.

This chapter is organized in six sections. In section 1.2, some more information and background about visa programs in the US, especially H-1B will be given. Section 1.3 reviews the literature of the economics of immigration. Section 1.4 describes the data used for the study along with some descriptive statistics. In section 1.5, the methodological approach will be explained, and the results will be discussed, and finally, section 1.6 will be about conclusion.

1.2 Background and history of US visas

In the NSCG data, as mentioned earlier, each individual clarifies if he/she is a native citizen, naturalized citizen, permanent resident (Green Card holder), or holding a visa. If a person holds a visa, the type of visa will be mentioned as well in four different categories: temporary work visa, study or training visa, dependent visa, or other.

Based on the definitions of the State Department of the United States, in total there are 10 different work visa categories[1]. Amongst all, H-1B program is the largest.

[1] See Table 1.1 for general descriptions.

Table 1.1: US work visa categories

Visa category	General description
H-1B: Person in Specialty Occupation	To work in a specialty occupation. Requires a higher education degree or its equivalent. Includes fashion models of distinguished merit and ability and government-to-government research and development, or co-production projects administered by the Department of Defense.
H-2A: Temporary Agricultural Worker	For temporary or seasonal agricultural work. Limited to citizens or nationals of designated countries, with limited exceptions, if determined to be in the United States interest.
H-2B: Temporary Non-agricultural Worker	For temporary or seasonal non- agricultural work. Limited to citizens or nationals of designated countries, with limited exceptions, if determined to be in the United States interest.
H-3: Trainee or Special Education visitor	To receive training, other than graduate medical or academic, that is not available in the trainee's home country or practical training programs in the education of children with mental, physical, or emotional disabilities.
L: Intra-company Transferee	To work at a branch, parent, affiliate, or subsidiary of the current employer in a managerial or executive capacity, or in a position requiring specialized knowledge. Individual must have been employed by the same employer abroad continuously for one year within the three preceding years and in most cases must have a bachelor's degree or equivalent.
O: Individual with Extraordinary Ability or Achievement	For persons with extraordinary ability or achievement in the sciences, arts, education, business, athletics, or extraordinary recognized achievements in the motion picture and television fields, demonstrated by sustained national or international acclaim, to work in their field of expertise. Includes persons providing essential services in support of the above individual.
P-1: Individual or Team Athlete, or Member of an Entertainment Group	To perform at a specific athletic competition as an athlete or as a member of an entertainment group. Requires an internationally recognized level of sustained performance. Includes persons providing essential services in support of the above individual.
P-2: Artist or Entertainer (Individual or Group)	For performance under a reciprocal exchange program between an organization in the United States and an organization in another country. Includes persons providing essential services in support of the above individual.
P-3: Artist or Entertainer (Individual or Group)	To perform, teach or coach under a program that is culturally unique or a traditional ethnic, folk, cultural, musical, theatrical, or artistic performance or presentation. Includes persons providing essential services in support of the above individual.
Q-1: Participant in an International Cultural Exchange Program	For practical training and employment and for sharing of the history, culture, and traditions of your home country through participation in an international cultural exchange program.

Ref.: http://travel.state.gov/content/visas/english/employment/temporary.html

The H-1 temporary worker visa was first established in 1952 as part of the Immigration and Nationality Act. It was letting foreign workers “of distinguished merit and ability” fill some positions temporarily. H-1 program had no caps or conditions to protect US workers at the time. It only had a requirement that both the worker and the job should be temporary (this requirement was even removed in 1970). In 1990, the H-1 program was divided into two types: H-1A and H-1B. The H-1A was only to bring educated nurses in the US to fill the nursing shortage, and it was stopped in the 1995. H-1B was at the time for all non-nursing skilled occupations. It required at least a bachelor’s degree or equivalent (in the relevant field if possible) (Lofstrom and Hayes, 2011). The H-1B visa will be valid for three years and could be renewed for another three-year period just once unless some exceptions exist. For an H-1B visa to be approved, the United States Department of Labor must confirm that the applying foreign worker will not displace or adversely affect the wages or working conditions of US workers (Lofstrom and Hayes, 2011). They should also at least get the prevailing wage in their fields. H-1B holders could change their employers if the new employer does the sponsorship for a new H-1B visa. This causes some mobility and to some extent frees such workers from being stuck in one job. The big advantage of H-1B is that people while working with the visa could start an employment-based permanent residence visa (Green Card) process.

There is a cap set by congress each year that limits the number of H-1B’s that could be issued in that year. Congressional cap on H-1B has changed frequently during years, as concerns about advantages and disadvantages of bringing skilled workers influenced legislation. Until 1998 the cap of 65,000 H-1B visas per year was set and not changed. It changed to 115,000 for 1999 and 2000 and increased to 195,000 for the next three years. Then it was again decreased to 65,000, and it is still the case. However, almost every year the number of H-1B’s issued exceed the cap number since based on the current regulations employers that are government research institutions, universities, or other nonprofits are exempt from the cap, and also up to 20,000 H-1B visas beyond the cap are available to those foreign temporary workers who have earned a master’s degree or higher from a university inside the United States (Lofstrom and Hayes, 2011).

“Temporary work visa” category in the NSCG data also includes Canadians and Mexicans on TN visas. Applicants must have job offers in the United States from a list of occupations that in most cases require at least a bachelor’s degree (Hunt, 2011).

The second visa category in the NSCG is “Study or Training” (F-1, J-1, H-3, etc.). Most students studying for a degree at a college or university will get F-1 visas, unless they have certain types of graduate fellowship, usually foreign funded, in which case they hold J-1 visas. Postdoctoral fellows and holders of foreign medical degrees doing a medical residency in the United States in general hold J-1 visas. There are also provisions for firms to engage trainees on J-1 or H-3 visas (Hunt, 2011).

There is another visa category in the NSCG named “Dependent” (F-2, H-4, J-2, K-2, L-2, etc.). These dependents could be spouses and children of principal temporary visa holders. For instance, a K-2 holder is the child of a K-1 holder, who himself/herself is the fiance of a U.S. citizen. While spouses of J-1 and L visa holders are eligible to work, spouses of H-1 and F-1 visa holders are not (Hunt, 2011).

1.3 Literature Review

Immigration has been a controversial topic among economists for a long time. The issue almost disappeared in the 1960’s, but again it became interesting among researchers as immigrant inflows were significantly increased (Card, 2009).

In general, there are two main research streams which focus on issues related to immigration effects. The first approach investigates the entire positive and negative effects of immigration on different components of the destination economy and labor market such as wages, employment opportunities, job security, expenditures in social programs, etc. This approach sometimes also studies the changes in equilibriums and tries to find the new equilibria after certain amount of immigrants (with specific qualifications) enter the labor market of the destination country. Doing this type of research on immigration is more intriguing and common when the researcher studies immigrant friendly countries such as the United States, Canada, Australia, etc.

The other approach focuses on the earnings of immigrants. Studies are carried out to check for significant differences between incomes of immigrants and those of natives controlling for different socio-economic and demographic factors. There has always been a concern about immigrants being under-paid compared to natives. Some believe that the reason could be related to some legal limitations that some immigrants might have which enable employers to exploit them in some ways. However, not all immigrants have such limitations. The general belief is that when an immigrant migrates to the host country, some of his/her human capital gained in the home country which used to be rewarding in that labor market is not much relevant and useful any more in the new labor market. So, the person needs to gain new human capital such as language proficiency, education, experience, etc. which is appropriate for the destination market. For sure, with time the level of relevant human capital could increase in immigrants and they will probably get closer to natives and might even overcome natives in wages. That is why so many researchers also study the assimilation process of immigrants over their life-cycles.

In this section, I will mention some of the papers that made significant contributions to the literature of economics of immigration. First, I will start by introducing influential papers in the literature that focus on and study either or both of the two above mentioned topics using cross-sectional data of the United States and other countries. Then, I will go over some of the important papers that study assimilation process of immigrants in the host country's economy, but instead of using cross-sectional data use longitudinal datasets to avoid cohort-specific bias problem raised by Borjas (1985) in the process of studying assimilation process. Finally, I review some papers that focus on comparing immigrant and native earnings based on the residency status and visa types using cross-sectional data of different countries. In order to give a better feeling regarding the contribution of papers, in each category, papers are presented based on their chronological order.

Although so many people have worked on immigration related topics, Chiswick (1978) is one of the pioneers in the field. Many of the studies carried out afterwards were based on his methodology introduced in his 1978 paper. The famous Mincerian earnings function is engaged in the paper to examine the effect of foreign birth, length of time in the US (years since migration), and US citizenship on the earnings of white men born outside the United States. He uses the 1970 census cross-sectional data for his study. As explained in the paper, since in the 1970 data most of the foreign born are white, his analysis is restricted to whites to avoid a confounding of the effects of race and foreign origin on earnings. It is also limited to men due to the problem of estimating labor market experience for women. His results show that although immigrants initially earn less than the natives, their earnings go up more rapidly with US labor market experience, and after 10 to 15 years their earnings equal, and then exceed, that of the natives. Chiswick also concludes that earnings are unrelated to whether the foreign born immigrants are US citizens or not. Long (1980) uses the exact same dataset and the same model as Chiswick (1978)'s. The only difference is that he does not exclude females from his research. Since labor force participation of women may not be continuous over their life cycles and "age minus schooling minus 5" might be unreliable as a proxy for women's experience, in order to control for experience in the absence of work history, Long brings number of kids of a woman and her kids' age structures along with her marital status into the model, and concludes that the earnings of foreign born females are about 13% higher than their native counterparts.

Doubtful about Chiswick (1978)'s results, Borjas (1985) writes an influential paper to re-examine the empirical basis for two results found and reported in Chiswick's paper: the earnings of immigrants grow quickly as they assimilate into the United States; and this rapid growth results in immigrants overtaking natives in earnings within 10-15 years after immigration. Borjas uses both 1970 and 1980 US census datasets and again runs a Mincer-based wage equation on each of them separately. Instead of putting "years since immigration" in the model, he uses dummy variables for different time periods of immigration, and by comparing the two estimates he shows that the cross-section regressions commonly used in the literature confound the actual assimilation impact with possible quality differences among immigrant cohorts. Instead of the fast growth found by the cross-section studies like that of Chiswick (1978), the cohort analysis predicts relatively slow rates of earnings growth for most immigrant groups. Borjas suggests the reason could be that the quality of the immigrants who migrated to the US from 1970 to 1980 was decreased.

Borjas (1994) in another well cited paper studies three major issues in Economics of Immigration using United States 1970 and 1980 census data: First, how do immigrants perform in the host country's economy? Second, what impact do immigrants have on the employment opportunities of natives? Finally, which immigration policy is most beneficial to the host country? He checks for different immigrant cohorts and different generations and concludes: the relative skills of newer immigrants decreased over the post-war period; it is unlikely that new immigrants reach parity with the earnings of natives; immigration could be responsible for the decline in the earnings of unskilled natives during the 1980s; immigration policy matters since it can only let immigrants in who are more skilled, and less probably will participate in public assistance programs; and finally, there exists a strong correlation between the skills of immigrants and the skills of their descendants. So, he deduces that immigration has a long lasting effect on the host country's economy. In another paper, Borjas (1995b) uses a simple economic framework to show how natives benefit from immigration. He argues that natives benefit from immigration mainly because of production complementarities between immigrant workers and other factors of production. Borjas concludes that these gains could be increased significantly if the United States follows an immigration policy that attracts more skilled immigrants.

In a very well cited and comprehensive study published in the handbook of labor economics, Borjas (1999) surveys the economic analysis of immigration by investigating the determinants of the immigration decision by workers in source countries and the impact of that decision on the host country's labor market and on immigrants' wages. Borjas in this paper theoretically and empirically studies different aspects of immigration in a comprehensive way and also reviews and critiques the previous studies. He includes his own previous works and results on immigration as well. This paper could be used as a reference for further and deeper understanding of the economics of immigration.

Friedberg and Hunt (1995) review the empirical and theoretical studies done by different authors about different countries to see the impact of immigrants on host country wages, employment and growth. They conclude that despite the popular belief, the literature does not provide much evidence that immigrants have a large adverse impact on the wages and employment opportunities of the natives. In terms of growth, they conclude that theoretical literature on immigration and economic growth shows that the impact of immigrants on natives' income growth depends on the human capital levels of the immigrants, but empirical researches on this issue give conflicting answers.

Friedberg (2001) uses the natural experiment of 12% population increase in Israel between 1990 and 1994, and by employing Instrumental Variables estimation shows that despite the common belief; massive immigrations from Soviet Union did not have an adverse impact on native Israelis' earnings and employment opportunities.

Card (2001) studies the impacts of new immigration on occupation-specific labor market using the 1990 US census data and finds that immigrant inflows over the 1980's slightly reduced wages and employment rates of less skilled natives in traditional gateway cities. In another paper, Card (2005) focuses on the questions "does immigration reduce the labor market opportunities of less-skilled natives?" and "have immigrants who arrived after the 1965 Immigration Reform Act been successfully assimilated?" Card uses the 2000 US census data to answer the first question and March CPS data from 1995 to 2002 for the second one, and does not find enough evidence supporting that immigrants have a negative impact on less educated natives. On the question of assimilation he finds that few of immigrants who come to the US without completed high school education will ever catch up with the average earnings of natives. However, most of their American-born children will catch up with the children of natives. Card (2009) in another study presents an overview on the connection between immigration and wage inequality, focusing comparisons across US major cities. He concludes that within broad education classes, immigrant and native workers appear to be imperfect substitutes, with a large elasticity of substitution. So, immigration has slight effects on wage inequality among natives.

Abramitzky et al. (2012) study the assimilation of European immigrants in the United States labor market during the "Age of Mass Migration" (1850-1913) using a newly-assembled panel data, and show that the average immigrant did not experience a substantial earnings difference upon first arrival and also experienced occupational advancement at the same rate as natives. They also show that assimilation patterns vary across sending countries and persist in the second generation.

In one of his more recent studies, Borjas (2013) again examines the evolution of immigrant earnings in the US labor market using 1970-2010 Census datasets. His results show that there are cohort effects both in the level of earnings that means more recent cohorts generally have relatively lower entry wages and also in the rate of growth of earnings suggesting more recent cohorts have smaller rates of economic assimilation. A part of this slowdown in wage convergence is due to reduction in the rate of human capital accumulation especially English language proficiency in more recent cohorts. The English learning process is significantly slower for larger national origin groups, and the growth in the sizes of these groups accounts for about a quarter of the decline in the rates of human capital accumulation and economic assimilation.

Rodríguez-Planas and Vegas (2014) run a study for Spain as well using National Immigrant Survey (ENI-2007) to compare assimilation process of Moroccan immigrants with Ecuadorians and Romanians (the two other largest groups of migrants to Spain). Employing Heckman-corrected estimates, their results show that controlling for all socio-economic factors, Moroccans have higher wages at arrival and this differential does not decrease over time.

Many studies in the literature use longitudinal datasets to avoid the cohort problems which were raised by Borjas. In order to come up with a more reliable assimilation rate, using a longitudinal data will be helpful. For instance, Hu (2000) uses Health and Retirement Survey (HRS) longitudinal data source which is a longitudinal survey of the population born between 1931 and 1941. He finds that the rate of increase of immigrant earnings is overstated in census-based cross-sectional studies, and the gap between immigrant and native earnings for more recent arrival cohorts is bigger than what was previously found.

Chiswick et al. (2005b) develop a model of the occupational mobility of immigrants using data on males from the Longitudinal Survey of Immigrants to Australia. Their study shows that the initial occupational status of immigrants may be an unreliable approximation of their ultimate occupational achievement. Based on the results, immigrants with higher level of (transferable) skills who are economic immigrants not refugees or family immigrants seem to have the most successful occupational adjustment through time. Moreover, they conclude living in an immigrant/ethnic concentration area seems to improve the occupational status of immigrants. Same authors in another paper in the same year (Chiswick et al. (2005a)) study the determinants of the level and growth in earnings of adult male immigrants in their first 3.5 years in Australia using the same data. They find that that assimilation happens and using this data the cross-section provides a good estimate of the longitudinal assimilation of immigrants.

Beenstock et al. (2005) create a longitudinal dataset by matching immigrants in Israel's censuses for 1983 and 1995 and show that it does not support the immigrant assimilation hypothesis, which predicts that the earnings growth for immigrants should vary inversely with duration.

Izquierdo et al. (2009) also use a new panel dataset "Continuous Sample of Working Histories" to examine the earnings assimilation of immigrants in Spain. They show that immigrants reduce the wage gap compared to natives by 15% during the first 5-6 years after arrival, but the earnings gap will not be gone completely. Based on their results assimilation is faster for South American and European immigrants compared to Africans.

Hall et al. (2010) using the 1996-1999 and 2001-2003 panels of the Survey of Income and Program Participation (SIPP) estimate wage differences for four groups: documented Mexican immigrants, undocumented Mexican immigrants, American-born Mexican Americans and native non-Latino whites. Their results show that after controlling for other factors, there exists 8% and 4% wage differences between documented and undocumented Mexican immigrant men and women, respectively. They also find large differences in returns to human capital with undocumented Mexican immigrants having the lowest wage returns to human capital and having very slow wage growth over time.

Some papers are focused on comparing immigrant and native earnings based on the residency status and visa types. Mithas and Lucas Jr (2010) use data on skills and compensation of more than 50,000 IT professionals in the United States over the period 2000-2005 and find that after controlling for socio-economic factors, foreign IT professionals with H-1B or other work visas earn a salary premium when compared with IT professionals with US citizenship. They also find that the salary premiums changes in response to changes made to the annual caps on new H-1B visas.

Lofstrom and Hayes (2011) analyze earnings differences between H-1B visa holders and US born workers in Science, Technology, Engineering, and Math (STEM) occupations and come up with similar results. They Combine H-1B data from US Citizenship and Immigration Services (USCIS) and data from the 2009 American Community Survey (ACS) to do the analysis. Their results show that H-1B's appear to have higher earnings compared to natives in some key STEM occupations, including information technology.

Hunt (2011) uses the 2003 wave of National Survey of College Graduates (NSCG) dataset and examines how immigrants perform according to the type of visa on which they first entered the United States. Her analysis actually differs from the previous works in its use of the entry visa instead of the current visa. She finds that immigrants who entered on a student/trainee visa or a temporary work visa have a large advantage over native college graduates in wages, patenting, commercializing and licensing patents, and authoring books or papers for publication or presentation at major conferences. Based on her results, much of such an advantage is explained by immigrants' higher education and field of study. She also finds that immigrants who entered on a student/trainee visa or a temporary work visa are more likely than natives to start a successful company. She finds that immigrants who came as legal permanent residents perform similarly to natives, while those who arrived as dependents of temporary visa holders or on other temporary visas perform worse than natives.

1.4 Data and Descriptive Statistics

As mentioned earlier, the 2010 wave of the National Survey of College Graduates (NSCG) which is collected by the US Census Bureau under the auspices of the National Science Foundation (NSF) is used in this study. The National Survey of College Graduates is a longitudinal biennial survey that provides data on the nation's college graduates, with particular focus on those in the science and engineering workforce. The program has been conducted since the 1970's. The survey samples are individuals living in the United States during the survey reference week, have at least a bachelor's degree in any field and are under the age of 76. This survey is a unique source for examining various characteristics of college-educated individuals, including occupation, salary, the field of study of the highest degree, the type of entry visa for immigrants and their current visa (described in section 1.2), whether each degree was received in the US, and demographic information. The 2010 wave of NSCG selected a part of its sample from the 2009 American Community Survey (ACS) respondents who indicated that they had a bachelor's degree or higher in any field of study. The remaining portion of the 2010 NSCG sample was selected from respondents to the 2008 NSCG survey[2].

The 2010 survey cycle in total includes 77,188 individuals. For my study, I only keep those who are living and working inside the United States during the survey reference week and are 65 years old or younger. 59,705 individuals that pass the criteria mentioned above remain in the dataset. In order to do the study, hourly wage of each individual is derived using annual salary, number of weeks worked per year, and number of hours worked per week. It seems that some respondents have confused annual weeks and months, or weekly and daily hours or for any other reasons have reported too big or too small annual salaries. Based on the 2013 report of the Labor Department of the United States, less than about 2% of annual wages in 2013 are more than 200,000 USD, while in the 2010 data more than 3% of people have reported bigger annual wages and about the same percentage of respondents have reported zero or very low incomes. Hence, I just keep those individuals who earn more than or equal to the 2010 federal minimum wage (7.25 USD) and less than 100 dollars per hour. After dropping those who do not meet such criteria and also those who show some other kinds of conflict in their responses, the number of individuals in the sample will decline to 54,813 out of which 31,260 are men and 23,553 are women. The (weighted) share of natives is 86.85%; naturalized US citizens have about 8.12% share; permanent residents constitute 3.35% of the (weighted) sample, and the remaining 1.68% goes to the temporary residents of the US. Specifically, the (weighted) shares of different visa categories are as follows: Temporary work visa 1.28%, Study or training visa 0.27%, Dependent visa 0.09%, and Other 0.05%. This means that 61.72% of immigrants (non-natives) in the weighted sample of 2010 are naturalized citizens, 25.5% of them hold green cards, 9.72% of them are with temporary

[2]Ref.: http://www.nsf.gov/statistics/srvygrads/#sd

work visas, 2.04% of them are on student or training visas, 0.66% of them are on dependent visa, and 0.36% of them are with other types of visas. Since people with dependent or other visas in the sample make very small groups, in my categorization, I consider both groups together as "Other". So, there will be 6 groups of people in my sample: Natives, Naturalized citizens, Permanent residents, Individuals on temporary work visa, Individuals on study or training visa, and Individuals on other types of visas.

Tables 1.2 and 1.3 show some descriptive statistics about some of the most important variables used in the analysis using the weighted sample (and sub-samples). As could be seen in table 1.2, immigrants on average have a larger hourly wage compared to natives, while the average ages of immigrants and natives are quite similar. The kernel density plots in graph 1 of figure 1.1 shows that immigrants' and native's wages are distributed similarly, so there will be no need to extend the wage analysis beyond mean regressions. According to table 1.2, inside the group of immigrants, naturalized citizens have the highest hourly wage which is probably because of four reasons: their mean of age is higher; on average they have bigger years since migration (YSM); they migrated to the US when they were younger, and they have the citizenship which might be advantageous in the US labor market. The second highest hourly wage which is not much less than that of naturalized immigrants is for persons with work visa. As mentioned earlier, H-1B program is the most important part of work visas and in order to issue an H-1B, there should be evidence that the person is highly skilled and no American qualified enough is available for the job. So, it makes sense if people on work visas get paid higher than most of other groups. Those with study/training visas are interestingly getting the minimum hourly wages compared to other groups. The reason for such difference could be that since people on student visas like F-1 are only eligible to work for their schools or some limited places, there will be some kind of the monopsony problem and they might be exploited by their employers. At the same time, since they are studying or on training programs, they just can have part-time jobs. Moreover, they are on average the youngest group with the least years since migration. These factors could also be some other reasons for their lower average wages.

Based on table 1.2, Naturalized immigrants are on average the oldest and people on student or training visa are the youngest group followed by individuals on working visa. As shown, after naturalized citizens, understandably, those on student/training visa have the lower mean of age at arrival. Individuals with work visa came to the US when they were averagely about 30 years old. Looking at the number of years since migration to the United States, a logical pattern could be seen. Naturalized citizens on average have been living in the US for about 24 years; permanent residents' (weighted) mean for years since migration (YSM) is about 12 years; people working with work visa have been in the US for about 7 years, and students or trainees have averagely past 4.2 years since the first time they entered the US.

Table 1.3 depicts the (weighted) means of education levels in different samples along with means of some other covariates used in the analysis. As shown, education level is higher in immigrants compared to natives. Both natives and immigrants mostly hold bachelor's degrees, but immigrants have fewer bachelor's degrees and more master's and doctorate degrees in return. Expectedly, study/training visa group has the lowest number of bachelors and highest masters and doctorates compared to other groups. The work visa group is ranked second for the highest post graduate degrees and the least under graduate degrees. This group of people, obviously, is supposed to be educated enough and well trained to get the high-skill jobs. Permanent residents group has a good number of post-graduates, as well. Many of members of this group could have gotten their green cards after being on H-1B program for a while. That could explain in part for higher education level.

Around 54% of all of the immigrants have earned their highest degrees from a university inside the United States. A big percentage of naturalized citizens of the sample have gotten their highest degrees from inside the US while just about one third of green card holders or people on work visa have their highest degrees from a school inside the United States. Surprisingly, about 45% of individuals studying or on training have earned their highest degrees from some universities inside the US.

Another interesting piece of information is about the (weighted) share of men in the samples and sub-samples. The whole sample and the group of natives have fair almost half and half shares of men and women. About 57% of the immigrants in the sample are men and this goes up to more than 60% in permanent residents, work visa, and study/training visa groups. This shows that immigration in the three latter groups of the sample is more done by male individuals, and so it makes sense that most of the people in the "Other visa" group which includes spouses and dependents of main immigrants are women. So many of people who migrate to the US on a work visa or get one after getting a degree are men.

Taking a look at the percentage of individuals with physical disabilities is also interesting. Although 7.75% of natives of the (weighted) sample have some kind of physical disabilities, this ratio is less in immigrants. Especially, the last three groups have a percentage about 3-4

Table 1.2: Weighted means of hourly wage and some covariates by residency status

	Hourly Wage (USD)		**Age**		**Age at Arrival**		**YSM***		**No. of Obs.****
	mean	S.D.	**mean**	S.D.	**mean**	S.D.	**mean**	S.D.	
Whole sample	31.13	(17.02)	43.09	(11.19)	-	-	-	-	54,813
Natives	30.79	(16.88)	43.08	(11.33)	-	-	-	-	42,889
Immigrants	33.37	(17.73)	43.13	(10.23)	23.24	(11.83)	18.68	(12.69)	11,924
Naturalized citizens	34.97	(17.95)	45.45	(10.18)	19.97	(11.98)	23.92	(12.18)	8,137
Green card	30.95	(18.08)	41.13	(8.76)	28.25	(10.26)	11.93	(8.55)	2,425
Work visa	33.71	(14.31)	36.32	(8.64)	29.38	(7.70)	6.94	(5.08)	953
Study/training visa	19.91	(10.89)	30.00	(4.82)	25.77	(4.54)	4.23	(3.63)	359
Other	21.20	(12.69)	43.90	(10.67)	32.27	(8.98)	11.63	(11.45)	50

* Years Since Migration **Number of Observations (not weighted)

Table 1.3: Weighted means of education levels and other covariates by residency status (in percent)

	Highest degree				**Highest degree from US**	**Male**	**Physical disability**
	Bachelor's	**Master's**	**Doctorate**	**Professional**			
Whole sample	63.37	27.37	3.53	5.73	93.71	50.39	7.58
Natives	64.53	26.92	2.77	5.77	99.67	49.39	7.75
Immigrants	55.67	30.34	8.52	5.48	54.35	56.98	6.50
Naturalized citizens	58.03	29.05	6.55	6.37	67.15	55.09	7.11
Green card	54.24	29.12	12.17	4.48	33.19	60.59	6.63
Work visa	49.85	36.81	10.58	2.76	34.81	61.93	3.13
Study/training visa	30.89	50.94	15.91	2.26	45.60	62.25	4.00
Other	53.34	35.84	2.33	8.48	12.51	23.57	3.62

percent which shows that people with disability either cannot or do not want to migrate as high skilled workers or students/trainees. One reason could be that people with disabilities are not much demanded as skilled workers in the US labor market.

In table 1.4 the distribution of highest degrees of each group of the weighted sample in different major fields is shown. Computer and information technology related fields are much demanded these days in the US labor market. A big part of issued H-1B's is also going to people who have such degrees. Interestingly, only 3.24% of natives in the sample have their highest degrees in this field, while it is about 10.5% in immigrants in general and 18.6% in the work visa group which complies with the H-1B share. While Mathematics and Statistics related or Physics related majors do not have a big share in most of the groups, engineering and other science and engineering related fields have quite considerable shares. In general, natives are more interested in getting non-S&E degrees while immigrants' majority goes for S&E related degrees. Engineering majors and computer related fields together stand for about 40% of highest degrees of those with work visas, and about 43.5% of people on student or training visas have their highest degrees either in engineering or in biological and environmental fields.

Table 1.5 shows the distribution of countries of origin for each immigrant group. There are some interesting facts in the sample. For example, Indian born immigrants have the biggest share in all of the weighted samples and sub-samples of immigrants. Almost 42% of the people in the work visa group of the weighted sample are Indians. According to US Citizenship and Immigration Services (USCIS), in 2009, 48% of the issued H-1B's were for Indians. Also, a significant portion of immigrants in different groups especially students are China born. Canada, UK, Mexico, and Iran are also the place of birth of so many of immigrants. Countries with significant shares are listed in table 1.5 along with their shares in each group.

Graphs 2 and 3 of figure 1.1 show the age-earnings profiles of natives and immigrants, respectively. It could be seen that (weighted) hourly wages of natives go up slower, and has a lower maximum but stays on top for a longer time, while (weighted) hourly wages of immigrants inclines faster and reaches a higher point, but then starts falling.

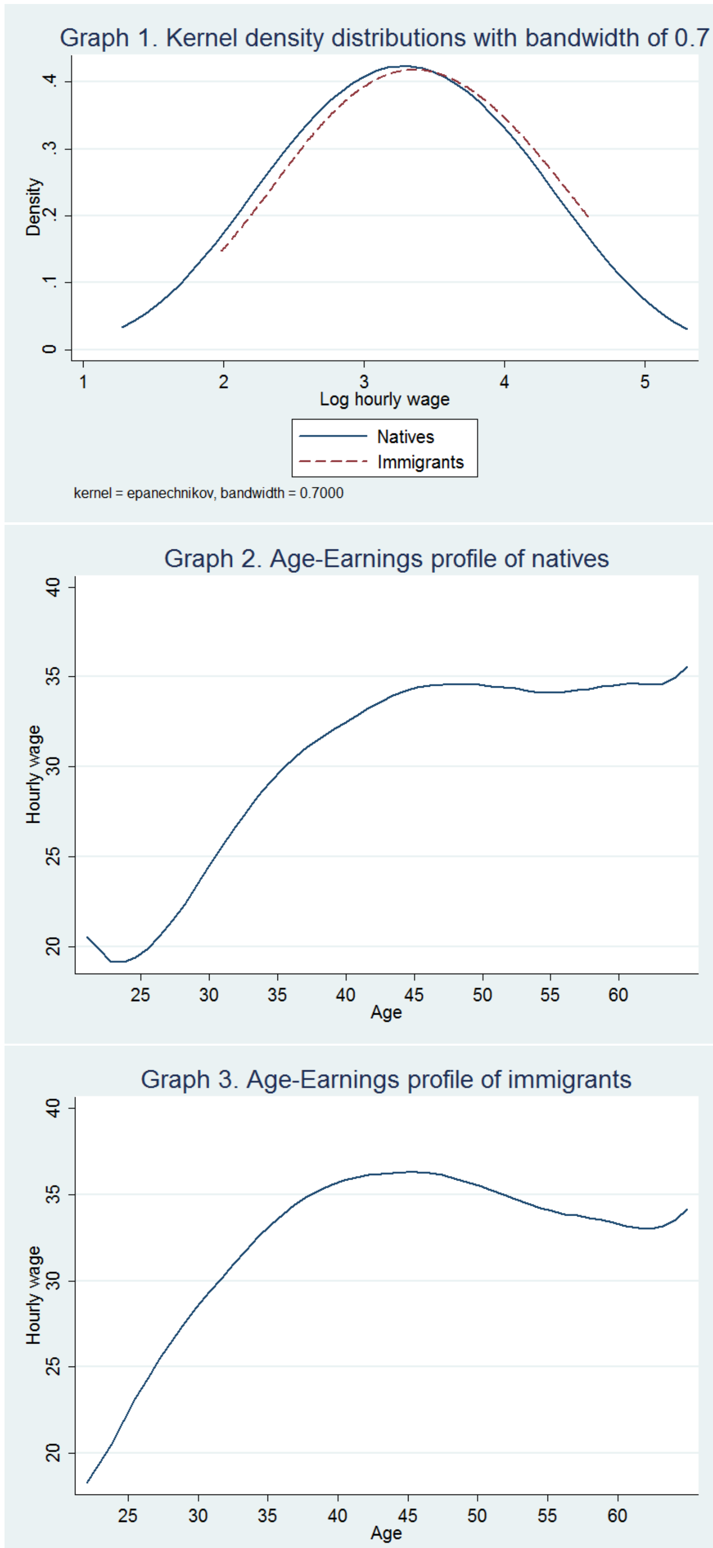
Table 1.4: Weighted means of field of study of highest degree by residency status (in percent)

	Computer & IT related	**Math., Stat., and OR.**	**Bio, Agri, and Env. sciences**	**Phys. and related sciences**	**Social and related sciences**	**Engineering (Computer and IT excluded)**	**Other S&E related**	**Non-S&E**
Whole sample	4.19	1.21	4.39	1.64	10.68	6.06	12.76	59.08
Natives	3.24	1.06	4.15	1.42	10.90	4.92	12.09	62.23
Immigrants	10.46	2.21	6.02	3.09	9.20	13.57	17.21	38.26
Naturalized citizens	9.23	2.31	5.32	2.98	9.67	11.96	18.36	40.16
Green card	10.86	2.14	6.88	3.35	9.19	13.66	16.92	37.00
Work visa	18.64	1.67	6.22	1.94	7.03	21.19	12.10	31.20
Study/training visa	7.36	2.96	16.57	9.16	6.70	26.96	11.56	18.73
Other	2.59	1.30	3.96	2.50	6.23	8.78	14.29	60.36

Table 1.5: Weighted means of place of birth by immigration status (in percent)

	India	**China**	**Canada**	**UK**	**Mexico**	**Iran**	**Other**
Immigrants	14.70	5.33	4.69	3.12	4.26	1.62	66.28
Naturalized citizens	11.15	5.09	3.78	2.47	5.00	2.27	70.24
Green card	12.35	5.02	7.46	5.42	3.30	0.67	65.78
Work visa	41.83	4.95	3.62	1.99	3.15	0.18	44.28
Study/ training visa	22.30	19.63	3.88	0.98	1.21	1.67	50.33
Other	14.64	2.98	1.77	0.08	0	0	80.53

Figure 1.1: Kernel density distributions and age-earnings profiles of natives and immigrants



1.5 Methodology and Results

The outcome of interest in this paper is log of hourly wage. I estimate a least squares model weighted with sample weights with robust standard errors as follows:

$$\log\, w_i = \beta_0 + I_i\boldsymbol{\beta}_1 + X_i\boldsymbol{\beta}_2 + \varepsilon_i \tag{1.1}$$

where i is the index for individuals in the equation 1.1 and I is a matrix containing dummies for all immigration and residency statuses such as different visa groups, permanent residency, and citizenship by naturalization. The dependent variable of the model as stated above is the log of hourly wage and X is a matrix which includes the socio-economic and demographic attributes of individuals in the sample such as the highest degree, whether the highest degree was earned in the United States, field of study of highest degree, age, immigrant's age at arrival in the US, years since migration, foreign and U.S. potential experience, arrival cohorts, sex, race, employment sector, self-employment status, firm size, and physical disability indicator. ε is the error term.

As mentioned earlier, on the right hand side of the model, there is a zero/one dummy for each immigrant status: naturalized citizen, permanent resident (Green Card), temporary work visa, temporary student/training visa, and other types of visas as the residual. Obviously, one and only one of these dummies will take 1 for each immigrant individual in the sample under study and they will all take zeros when it comes to a native. Hence, the estimated coefficient on the indicator (dummy) variable of any above mentioned immigrant status (if significant) represents the log earnings difference between US born (native) workers and that group of immigrants.

Table 1.6 presents the results of the OLS models ran in this study. Column (1) reports the results of regressing log of hourly wage on the above mentioned 5 immigrant groups with no further variable on the right-hand side (as mentioned earlier the dummy for natives is omitted to find out about log earnings differences). The estimated coefficients of work visa and study/training visa dummies are reported in the table. Column (1) shows that unconditionally, individuals on temporary work visas earn 15 log points (15%) more than natives.

In the second regression, field of the highest degree is controlled for in 8 major groups: "Computer and IT related fields including computer engineering and IT engineering", "Mathematics, Statistics and operations research", "Biology, agricultural and environmental sciences",

Table 1.6: OLS Regressions Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Work visa	0.15	0.04	0.08	0.18	0.29	0.26	0.25	0.22
	(5.28)	(3.17)	(2.32)	(5.20)	(6.60)	(3.19)	(3.10)	(2.73)
Study visa	-0.42	-0.53	-0.55	-0.31	-0.26	-0.27	-0.27	-0.17
	(-8.58)	(-11.78)	(-10.83)	(-5.64)	(-4.16)	(-3.05)	(-3.04)	(-1.86)
Highest degree								
Master's	-	-	0.20	0.17	0.22	0.22	0.22	0.24
			(15.90)	(13.89)	(17.48)	(17.57)	(17.35)	(19.50)
Doctorate	-	-	0.35	0.29	0.36	0.34	0.33	0.37
			(22.01)	(17.83)	(20.37)	(19.31)	(18.44)	(20.70)
Professional	-	-	0.49	0.47	0.49	0.47	0.46	0.47
			(22.71)	(22.95)	(23.70)	(22.64)	(22.25)	(22.34)
Age	-	-	-	0.18	0.16	0.15	0.14	0.15
				(7.30)	(6.64)	(6.04)	(5.93)	(6.39)
Field of Study								
Computer & IT	-	0.32	0.34	0.34	0.34	0.30	0.29	0.23
		(19.83)	(21.32)	(21.88)	(22.13)	(18.68)	(18.11)	(14.18)
Math, Stat, and OR	-	0.22	0.24	0.24	0.21	0.19	0.18	0.17
		(7.71)	(8.92)	(9.27)	(8.18)	(7.32)	(7.14)	(7.40)
Bio, Agri, and Env.	-	0.01	0.02	0.04	0.02	0.01	0.00	-0.02
		(0.51)	(1.42)	(2.47)	(0.98)	(0.41)	(0.17)	(-1.14)
Physics and related	-	0.21	0.20	0.18	0.16	0.12	0.12	0.09
		(9.82)	(9.42)	(9.38)	(8.24)	(6.50)	(6.32)	(4.57)
Social and related	-	0.01	0.05	0.06	0.05	0.05	0.05	0.03
		(0.68)	(3.26)	(4.47)	(3.71)	(3.95)	(3.87)	(2.26)
Engineering	-	0.35	0.38	0.37	0.36	0.31	0.30	0.25
(w/o Comp. and IT)		(26.55)	(29.45)	(29.78)	(28.74)	(23.26)	(22.87)	(19.27)
Other S&E related	-	0.25	0.20	0.19	0.19	0.21	0.20	0.18
		(19.34)	(16.00)	(15.69)	(15.71)	(17.10)	(16.77)	(15.21)
Male	-	-	-	-	-	0.13	0.13	0.11
						(11.87)	(11.95)	(10.30)
Phys. dis.	-	-	-	-	-	-	-0.15	-0.13
							(-7.38)	(-6.42)
English spk. countries	-	-	-	-	-	0.06	0.08	0.07
						(1.81)	(2.25)	(2.12)
R^2	0.008	0.060	0.130	0.197	0.214	0.230	0.238	0.290

Note.—Coefficients from OLS regressions using 54,813 observations, weighted with survey weights, with log hourly wage as the dependent variable. Robust *t*-statistics are reported in parentheses. The omitted category is US natives. The estimates of naturalized citizen dummy, permanent resident dummy and "other visas" dummy are not reported. The other variables which are controlled for but not reported are US/non-US degree dummy (present in regressions 3 to 8) age^2 and age^3 (reg. 4 to 8), age group at arrival dummies and potential foreign and US experience (reg. 5 to 8), years since migration to the US with a cubic, the arrival cohort (decade) dummy, and dummies for being from India, China, Canada, UK, Mexico, and Iran (reg. 6 to 8), race dummies (regressions 7 and 8), and employer sector dummies, employer size dummies, employer region dummies, and self-employed dummy (present only in the last regression).

"Physics and related sciences", "Social and related sciences", "Engineering excluding computer and IT engineering", "other science and engineering related fields", and finally "non-science and engineering related fields". The latter group is kept out as the base group. As shown in the table, only controlling for field of study is sufficient to reduce the premium to only 4%. People with work visa earn only 4% more than natives with a similar field of study. That makes sense since as showed in the previous section, most of the immigrants especially those with work visas have their highest degrees in computer related, engineering, or other S&E related fields which are paid better in the labor market while majority of natives have degrees in social sciences or non-S&E fields.

In the third regression the highest degree (bachelor's, master's, doctorate, and professional) and whether the highest degree is earned inside the United States are also controlled for (bachelor's degree is chosen as the base group). The coefficient on the work visa dummy in such a case represents the difference between the average native and the average immigrant with a work visa who obtained a US degree the same as that of a native in level and field. Such an immigrant has about 8% premium compared to the native.

In the fourth regression, I control for cubic in age (immigrants with work visa are on average about 7 years younger than natives). The estimated coefficient on the work visa dummy now represents the difference between natives and work visa holders with the same ages who both received their highest degrees in the US similar in level and field. As seen in the table, controlling for age along with all the previous variables, an immigrant with a temporary work visa on average will get 18% more than the average native.

In the fifth regression, I control for the age group at arrival in the United States (10, 20, 30, 40, 50, and 50+), and also for potential US and foreign experience[3]. Since these variables are correlated to each other, controlling for one without the others could give misleading results. The estimated coefficient on the work visa dummy now represents the difference between natives and work visa holders who arrived in the United States after birth, received the highest degree in the US similar in level and field to that of natives, and only has US experience. Like having US education, US experience will cause a premium and younger

[3]A technique similar to what used in Hunt (2011) is engaged to calculate any potential US or non-US experience. Potential U.S. experience is calculated as years since earning the highest degree if the degree was gotten in the United States or by a native, or as years since migration if the highest degree was obtained abroad by an immigrant. Potential foreign experience is defined as the difference between the year of arrival in the US and the year of getting the highest degree for immigrants who received their highest degrees abroad and zero for others.

arrivals will have a positive effect on wage. As depicted in the table, controlling for all the variables above, an immigrant with a temporary work visa on average will get 29% more than the average native.

In the sixth run, a cubic in number of years since migration to the US will be controlled for along with gender, the arrival cohort (decade), whether the person is born in an English speaking country[4] as a native English speaker, and whether the person is from India, China, Canada, UK, Mexico, or Iran as countries that have significant shares in the whole population of immigrants in the United States. People with English mother tongues will certainly have better chances in the US labor market and it will be an advantage in terms of human capital for native English speakers. The results of the regression also confirm this fact. Results show that being a man has also a positive more than 10% effect on the wage which could be because of strength, type of education, or other reasons. After controlling for all these parameters, column (6) shows a 26 log point difference against natives. Results of this regression also shows a significant (about 20%) wage premium for people who are born in the UK compared to the other people, while this number is about 8% for people born in India or China compared to people born elsewhere. Other than being a native English speaker, one of the reasons that people born in UK have such a considerable premium over other people could be the similarities of UK labor market and US labor market. It might be the case that, in both labor markets similar human capitals are appreciated.

In column (7), I add controls for different races and also for any physical disability. The races are categorized into seven different groups: Asian, American Indian/Alaska native, Black, Hispanic, White, Pacific islander, and multiple races. Asians are kept out as the base group. American Indian/Alaska native, Black, Hispanic, and White races show a negative premium in wages in the results compared to Asians and the relative estimated negative premiums for the first three groups are 28%, 9%, and 12% lower than Asian race, respectively. After controlling for race and disabilities other than the previously mentioned factors, column (7) reports a 25% difference between natives and immigrants working in the US with a temporary work visa.

Finally, in the last regression I control for employer sector, size, and region along with whether an individual is self-employed. United States is categorized into 9 different regions in the

[4]Countries are Canada, Bermuda, Jamaica, Antigua-Barbuda, Bahamas, Barbados, Dominica, Grenada, St. Vincent, Trinidad and Tobago, Guyana, the United Kingdom, Ireland, Northern Ireland, Liberia, South Africa, Australia, and New Zealand (Borjas, 2013).

NSCG data: New England, Middle Atlantic, East north central, West north central, South Atlantic, East south central, West south central, Mountain, and Pacific. The first region is chosen as the base, and most of the regions above show a negative wage premium compared to the New England region. However, as discussed in Hunt (2011), immigrants mostly live disproportionately in high-wage regions. So, region controls are appropriate if they pick up difference in price levels but inappropriate if they pick up genuine regional productivity differences of college graduates. College graduates in California, for instance, might be more productive than college graduates elsewhere and so on. Sectors are divided into 3 major groups: Educational, Governmental, and Business/Industry. I chose the first category as the base groups and the second two sectors show 12 and 17 percent wage premiums compared to the educational sector, respectively. Employer size is known by its number of employees. NSCG has grouped employers into 8 groups: "with 10 or fewer employees", "with 11-24 employees", "with 25-99 employees", "with 100-499 employees", "with 500-999 employees", "with 1000-4999 employees", "with 5000-24999 employees", and "with 25000+ employees". I took the smallest size as the base and the results strongly show that the bigger the firms are in number of employees, the bigger the wage premium of their employees will be. Results show that the employees of the last group on average get 30% more than those of the smallest size. One reason probably is that bigger companies are more famous and successful in their own fields and they hire more skilled, more educated, and more experienced individuals. As shown in column (8), after controlling for all socio-economic and demographic characteristics available in the data, individuals on a work visa are still getting 22% more than natives.

The reason for such a phenomenon could be that people who hold work visas are selected individuals who earned the work visa in a tough competition as they were capable people with high abilities. There are for sure some unobserved characteristics such as ability, motivation, etc. that are highly correlated with earnings and are not controlled for in the model. Those are causing such a significant wage difference between natives and those immigrants who hold work visas.

As reported in all of the columns of table 1.6, those immigrants who are on temporary study/training visa have a substantial negative premium compared to natives. The difference changes with different controls, but in all the 8 regressions it shows a significant negative difference. That phenomenon could be explained by taking into consideration that people on F-1, J-1 or other temporary study/training visas are in the US to study or do research and they mostly are only eligible to work for the school they are studying at or some other handful of places. So, first of all the monopsony problem could be the case, and secondly, since the main duty of such people is studying, those who work at the same time will do part-time jobs, and the payments for part-time jobs are in general less than full-time careers.

Results of table 1.6 show robust results for different highest degrees and different fields of study. As shown in the table, people with master's degrees on average get 20-25% more hourly wages compared to those who hold bachelor's degree; those holding doctorate degrees get about 35% more, and owners of professional degrees have about 45-50% wage premium compared to those with bachelor's degrees. Also, people with the highest degree in computer and IT related fields on average earn 20-30% more than average people with non-S&E degrees; Mathematics, Statistics, and OR related majors get about 20% more; physics and related majors have about 10% premium; Social sciences and related field get almost similar; while engineering fields (excluding IT and computer) and other S&E related field on average have about 25 and 20 percent premium compared to non-S&E fields, respectively.

In this study, I also controlled for interaction of male dummy and work visa dummy, interactions of dummies for country of birth (English speaking countries and those mentioned above with big shares of immigrants in the US) and work visa, and interactions of dummies for field of study of highest degree and work visa to see if work visa premium is different for men and women, people with different fields of study, and people from different countries. I ran different regressions to check for interaction terms. The previously reported results are robust to including the interaction terms. Table 1.7 shows the results of some covariates with significant estimates including the interaction terms for some of the regressions. None of the regressions which controlled for interactions of fields of study and work visa estimated significant coefficients for any of the fields interacting with work visa. So, there is no evidence that people with different fields of education and specialty who are on a temporary work visa have different premiums, and consequently those results are not reported in the table 1.7.

In column (1) of table 1.7, I control for the interaction of male dummy and work visa dummy variable other than all the covariates used in the last regression reported in table 1.6. In column (2) the interaction of English speaking countries dummy and work visa dummy is controlled for instead of the interaction of male dummy and work visa dummy. In the third regression which is reported in column (3) instead of the last two interactions in (1) and (2), I controlled for the interactions of dummies for country of birth (India, China, Canada, UK, Mexico, and Iran) and work visa. Column (4) reports the results of a regression in which

Table 1.7: OLS Regressions Results (interactions)

	(1)	(2)	(3)	(4)	(5)	(6)
Work visa	0.30	0.20	0.24	0.28	0.30	0.30
	(3.33)	(2.51)	(2.83)	(3.10)	(3.25)	(3.21)
Male * work visa	-0.13	-	-	-0.12	-0.12	-0.12
	(-2.29)			(-2.19)	(-2.26)	(-2.24)
English spk. countries * work visa	-	0.19	-	0.17	-	-0.05
		(1.85)		(1.76)		(-0.44)
Iranian * work visa	-	-	-0.39	-	-0.45	-0.45
			(-4.23)		(-4.57)	(-4.53)
Chinese * work visa	-	-	-0.21	-	-0.19	-0.19
			(-2.39)		(-2.16)	(-2.15)
Canadian * work visa	-	-	0.28	-	0.27	0.32
			(1.68)		(1.80)	(1.83)
Indian * work visa	-	-	-0.06	-	-0.02	-0.03
			(-0.85)		(-0.38)	(-0.40)
British * work visa	-	-	0.09	-	0.13	0.18
			(0.56)		(0.85)	(0.98)
Mexican *work visa	-	-	-0.11	-	-0.08	0.08
			(-0.78)		(-0.56)	(-0.57)
$\mathbf{R^2}$	0.290	0.290	0.290	0.290	0.290	0.290

Note.—Coefficients from OLS regressions using 54,813 observations, weighted with survey weights, with log hourly wage as the dependent variable. Robust *t*-statistics are reported in parentheses. The omitted category in all the 6 regressions is US natives. The estimates of naturalized citizen dummy, permanent resident dummy and "other visas" dummy are not reported. The other variables which are controlled for in all of the 6 regressions but not reported are dummies for fields of study of the highest degree, highest degree dummies, US/non-US degree dummy, age with a cubic, age group at arrival dummies, potential foreign and US experience, years since migration to the US with a cubic, gender dummy, the arrival cohort (decade) dummy, native English speaker dummy, dummies for being from India, China, Canada, UK, Mexico, and Iran, race dummies, physical disability dummy, employer sector dummies, employer size dummies, employer region dummies, and self-employed dummy.

Table 1.8: OLS Regressions Results (2003 data)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Work visa	0.07	-0.02	-0.08	0.07	0.14	0.13	0.14	0.14
	(2.93)	(-0.87)	(-3.45)	(2.77)	(4.47)	(2.22)	(2.49)	(2.45)
Study visa	-0.49	-0.56	-0.66	-0.46	-0.42	-0.41	-0.40	-0.31
	(-10.42)	(-12.31)	(-15.06)	(-10.36)	(-8.98)	(-6.25)	(-6.02)	(-4.63)
$\mathbf{R^2}$	0.003	0.039	0.092	0.118	0.129	0.151	0.154	0.205

Note.—Coefficients from OLS regressions using 76,423 observations, weighted with survey weights, with log hourly wage as the dependent variable. Robust *t*-statistics are reported in parentheses. The omitted category is US natives. The estimates of naturalized citizen dummy, permanent resident dummy and "other visas" dummy are not reported. The other variables which are controlled for but not reported are dummies for fields of study of the highest degree (present in regressions 2 to 8), dummies for highest degree and US/non-US degree dummy (reg. 3 to 8) age with a cubic (reg. 4 to 8), age group at arrival dummies and potential foreign and US experience (reg. 5 to 8), years since migration to the US with a cubic, gender dummy, the arrival cohort (decade) dummy, native English speaker dummy, and dummies for being from India, China, Canada, UK, Mexico, and Iran (reg. 6 to 8), race dummies and physical disability dummy (regressions 7 and 8), and employer sector dummies, employer size dummies, employer region dummies, and self-employed dummy (present only in the last regression).

both first two interactions are included, and column (5), instead, controls for the interactions of dummies for country of birth and work visa and the interaction of male dummy and work visa dummy. Finally, the last regression (column (6)) includes all of the three types of interactions.

As can be seen in table 1.7, men on work visa have smaller wage premiums compared to women on work visas; people with work visas who are born in English speaking countries and are native English speakers have higher premiums compared to people who hold work visas and are from other countries with non-English mother tongues; and Iranians and Chinese who hold work visas such as H-1B, have much lower premiums compared to others while Canadians working in the US with a working visa have a higher premium compared to others from other countries with work visa. These are interesting differences to know, but different unobserved and uncontrolled reasons such as discrimination, feminism, or even racism could be causing the results reported in table 1.7.

As mentioned earlier, the other goal of this study is to compare the wage premiums for immigrants on a work visa in 2003 and 2010 using the 2003 and 2010 waves of the NSCG dataset. So, I also ran the same regressions using 2003 data and compared the results for the work visa dummy. Table 1.8 reports the results of running my regressions on 2003 data.

As can be seen in tables 1.6 and 1.8, all the eight regressions show at least 6% difference in the wage premiums of the people holding temporary work visas comparing years 2003 and 2010. Table 1.9 reports the summary statistics of 2003 and 2010 (weighted) samples for natives and immigrants with work visa. As reported in the table, the averages of age and education have not changed significantly in natives comparing the two years. Comparing the other group (work visa) in 2003 and 2010, however, shows some differences: In both years immigrants on work visa have almost the same mean of age, while the averages of years since migration and the education level are slightly more in 2003. The share of non-S&E majors among work visa holders is also less in 2003. Although most of the human capital factors are on average better in work visa holders of 2003 compared to those of 2010, the estimated wage premium for this group over natives was lower in 2003 and with time this premium is increased to something significantly higher in 2010.

One of the very important differences between these two years that could probably explain this increase in the wage premium is that, as mentioned earlier, in 2003 the H-1B cap was 195,000, while it was only 65,000 (one third of 2003 cap) in 2010. Technology boom happened

Table 1.9: Weighted means of some covariates for natives and work visa

	2003		2010	
	Native	**Work visa**	**Native**	**Work visa**
Hourly wage (USD)	28.21	29.71	30.79	33.71
	(15.68)	(15.15)	(16.88)	(14.31)
Age	43.76	35.59	43.08	36.32
	(9.88)	(6.85)	(11.33)	(8.64)
Age at arrival	-	28.05	-	29.38
		(6.51)		(7.70)
Years Since Migration	-	7.54	-	6.94
		(4.47)		(5.08)
Highest degree:				
Bachelor's	64.22	40.87	64.53	49.85
Master's	26.44	36.87	26.92	36.81
Doctorate	3.11	16.04	2.77	10.58
Professional	6.22	6.21	5.77	2.76
Highest degree in US	-	43.25	-	34.81
Field of Study:				
Computer & IT	2.68	15.56	3.24	18.64
Math, Stat, and OR	1.35	3.50	1.06	1.67
Bio., Agri., and Env.	4.16	8.88	4.15	6.22
Physics and related	1.80	4.68	1.42	1.94
Social and related	10.36	8.17	10.90	7.03
Engineering (w/o Comp. and IT)	5.41	20.36	4.92	21.19
Other S&E related	12.46	14.99	12.09	12.10
Non-S&E	61.79	23.84	62.23	31.20
Phys. dis.	5.67	2.87	7.75	3.13

Note.—Except for hourly wage, age, age at arrival, and years since migration, all of the other weighted means are in percent. Numbers in parentheses show the standard deviation of each covariate.

in this period and the need for skilled labors in different fields especially in high-tech fields was increased during these years. United States' labor market did not have enough skilled professionals in it to satisfy the demand. American firms needed professionals from out of the US as skilled immigrants to reduce the shortage, but not only was not the H-1B cap increased, but also it was lowered to one third of what it was in 2003. The complicated and timely process of issuing a work visa in some cases also made immigration of the highly skilled labor slower. The high demand was not accompanied by a significant supply. So, the wages paid to experts on work visas got increased.

Another probable problem of the supply side that might be able to explain some part of the increase in the premium from 2003 to 2010 is the education system of the United States which (as shown in tables 1.3 and 1.4) is not motivating native students to go for more scientific and math based majors in higher degrees by focusing more on mathematics and other sciences in schools. Bringing professionals from outside of the United States could have some negative effects (as discussed in the introduction section) on the US labor market and those people already active in the market, but training natives to become professionals will be a good policy to meet the increasing demand of US firms with the least social cost.

1.6 Conclusion

It is certain that because of some sponsorship costs and some federal regulations entering the job market and getting hired is on average harder and more complicated for people with any visa compared to permanent residents (Green Card holders) or citizens. Employers obviously prefer to hire people who cost them less in terms of sponsorship or any other legal expenditure while qualified enough for the job. They also prefer natives or at least permanent residents since by hiring them (instead of foreign high-skilled workers with similar qualifications) they could count on longer cooperation with less legal difficulties. Yet, when a non-permanent resident individual gets a job and acquires a working visa after lots of competition and effort, the case will be different. Now, it is not for sure that some significant wage differences exist comparing those on work visas with natives. Some people believe that foreign skilled workers are a cheap source of labor for the United States. In other words, there is a general belief that there exists some wage discrimination against immigrants who are working with working permits (H-1B visa or other working visas) compared to the natives in the United States labor market.

The 2003 and 2010 waves of National Survey of College Graduates (NSCG) are used in this study to examine how well immigrants working in the United States with work visa and at least a bachelor's degree are getting paid compared to college graduate natives. In this study, I check for any significant wage disparities between the two groups controlling for socio-economic and demographic characteristics of individuals. Moreover, I study the change of such a wage premium (for or against immigrants holding work visas compared with natives) from 2003 to 2010. I also analyze to see whether this wage premium is different for men and women, for individuals originally from some specific countries which send or has sent a significant number of immigrants to the United States, or for people having different fields of study.

After controlling for most of the factors that are affecting earnings, I find that contrary to the common belief that foreign workers are cheap labor force, after controlling for socio-economic and demographic characteristics, skilled immigrants holding temporary work visas have a 22 log point wage premium over American natives. The results also show that the premium has even significantly increased from 2003 to 2010 (from 14% to 22%). According to results, this wage premium is less for men compared to women and different for individuals originally from some specific countries like Canada, China, and Iran, but there is no evidence supporting different premiums for people having different fields of study.

Chapter 2

Earnings Gap Between Highly Educated Immigrants and US-born Counterparts: Analysis of Dynamics

2.1 Introduction

The question of the earnings gap between immigrants and natives and also economic assimilation of immigrants goes back to more than 35 years ago, when the labor economists first started this line of research. A big part of this literature is studying the adjustment of immigrants in the labor market of the host country. Immigrants move to the host country with different skills and different packages of human capital. There is no guarantee that the human capital (mainly education and experience), that they have brought with themselves to the host country, matches the demanded human capital in the host country's labor market. So, logically, the longer they stay in the host country (assuming that their time in the host country is either spent on studying or on working), the more they acquire and accumulate the human capital which is appreciated by the host country's labor market. Cultural adjustments and enhancements in the local language proficiency will also happen with the more time immigrants live in the destination country[1]. Apparently, cultural adjustments (such as getting used to the common work ethic in the host country) and better language proficiency could positively affect immigrants' job market outcomes as well.

Chiswick (1978) was the first person who introduced the concept of "Years Since Migration" for immigrants, and brought it into the famous Mincer (1974) wage equation with the explanation that the effect of years spent on education or on gaining experience in the home country of immigrants could be different from the effect of time spent in the destination country. He, for the first time, used that concept to find the number of years needed for the earnings gap between immigrants and natives to be filled employing cross-sectional approach. After his 1978 paper got published, many researchers either criticized Chiswick's methodology and the correctness of his results, or used similar methodology and calculated the earnings gap between natives and immigrants and the assimilation rate for different population groups in different countries using various datasets. Most of these researchers use one or multiple cross-section(s) (depending on the availability of the data) and find an earnings gap against immigrants at entry time and a rate at which immigrants catch natives in wage and the gap is filled. It is argued in most of the cases that the earnings convergence is caused

[1] Cultural and language differences very much depend on immigrant's country of origin. Immigrants who come from western countries encounter less cultural differences than those who come from eastern ones. Similarly, depending on the mother tongue of immigrants, it might be harder for some groups to learn English than others. People who migrate to the United States from English speaking countries, especially, Canada, United Kingdom, Australia, New Zealand, and South Africa have an important advantage compared to other immigrants.

by the fact that immigrants are more able, more motivated, and/or more hardworking than natives. However, most of these characteristics cannot be observed and controlled for.

There have always been debates in the immigration literature on how to interpret findings gained from cross-sectional approach. Borjas (1985, 1995a) emphasizes that the cross-sectional approach might not measure the true rate of assimilation because there may be significant differences in earnings potentials between different immigrant year-of-arrival cohorts. In fact, Borjas' studies show the presence of some remarkable "cohort effects". He argues that more recently arrived immigrant cohorts might have lower or higher destination job-market-specific skills than earlier cohorts. Accounting for these cohort effects in wage levels might substantially change the rate of economic assimilation derived from cross-sectional approach (Borjas, 2013). In other words, instead of providing a measure of earnings assimilation, the cross-sectional results might actually capture changes in unmeasured dimensions of immigrants' skills in different arrival cohorts. In that case, the cross-sectional results will not have much to say about immigrant-native earnings convergence (Lemos, 2013).

Another type of bias, called "survivor bias", might also occur when multiple cross-sections from different years are used to find the assimilation or convergence rate. For instance, if return migration of immigrants who are less able and consequently less successful in the labor market happens, cross-sectional results will over-estimate the real assimilation rate. On the other hand, if more able immigrants return to their home countries or migrate to other countries that have more rewarding labor markets for their skills, the cross-sectional estimates will be under-estimating the real rate. A big part of the literature of the last 30 years deals with these different types of biases which could be caused by selectivity in the data. As recognized by different researchers, such as Chiswick (1980) and Borjas (1985), the ideal way of getting to accurate estimates and avoiding the above mentioned biases is using a balanced longitudinal (panel) data which follows the same people over time. However, that type of data is rare and hard to find.

To overcome cohort bias problem, Borjas (1985) proposed applying the "synthetic cohort methodology" (SCM) by using a succession of cross-sections to construct synthetic panel data. For instance, immigrants who migrated at the age of 20 in 1970 may be compared with immigrants aged 30 in 1980 who migrated in 1970, and so on. SCM does not compare the same immigrants because of out-migrations, deaths and sampling issues, but it does, at least partially, control for cohort effects. However, as Borjas (1985) himself was aware, the synthetic cohort methodology is not solving the survivor bias (Beenstock et al., 2005).

Other than the aforementioned concerns about the cross-section derived estimates, there are also some disputes about the definition of "economic assimilation". Beginning with Chiswick's (1978) paper, many studies implicitly or explicitly use a definition that defines the concept of economic assimilation to be the rate of wage convergence between immigrants and natives in the host country. LaLonde and Topel (1992) present a very different definition of the process: "assimilation occurs if, between two observationally equivalent (foreign-born) persons, the one with greater time in the United States typically earns more" (LaLonde and Topel, 1992). This definition only uses the immigrants' data to estimate the effect of years since migration. Hence the base group in the LaLonde-Topel definition of economic assimilation is the immigrant himself. These two definitions can lead to very different results. For instance, one can come up with a positive assimilation rate using the latter definition, while the former definition gives a negative result (Borjas, 1999). The present study uses the first (former) definition to calculate convergence/divergence rate between natives' and immigrants' wages.

My objective in this chapter is to estimate the earnings gap between highly educated natives and highly educated immigrants (based on current residency status of immigrants: naturalized citizens, permanent residents (Green Card holders), and visa holders) upon arrival, and also study the alterations of this gap with time using the cross-sectional and longitudinal approaches. Moreover, I check to see whether or not there exists any significant difference between the results achieved from the two approaches. I run the models on seven sub-sets of the sample: 1) whole sample, 2) men, 3) women, 4) natives plus immigrants who migrated to the US on an immigrant (permanent residence) visa, 5) natives plus immigrants who migrated to the US on a work visa, 6) natives plus immigrants who migrated to the US on a study visa, and 7) natives plus immigrants who migrated to the US on a dependent visa.

My results, in contrast to results of almost all of the previous studies, show that immigrants with a bachelor's degree or higher have a huge premium over their native counterparts. The more interesting result is that not only the wage gap at entry is positive for immigrants, but also this gap between natives and different groups of immigrants even gets wider for the first 10-20 years of immigrants' residence in the United States. Besides, my results gained from the two different approaches point to distinct routes that different groups and sub-groups of immigrants go from the time they enter the US: Cross-section results show a higher premium for all groups of immigrants at entry compared to what longitudinal results show. Convergence/divergence rates calculated for different groups using the two approaches are also different.

One of the contributions of this study to the literature is finding the earnings gaps and also assimilation rates for a highly educated population with science and engineering (S&E) degrees. Previous studies have used datasets that represent all groups of immigrants, most of whom have very low education levels compared to the average native. I also, for the first time, compare immigrants to natives based on their residency status, and also based on their first type of visa by which they migrated to the United States. Another advantage of this work compared to similar studies is employing a very rich data with detailed information about education, employment, and demographics of each individual. While education information is not available in many of the datasets, especially in administrative data, this data not only does give information about the level and major of the last five highest degrees of each individual in the sample, but also provides information about whether or not that degree is earned from a school inside the United States.

This chapter is organized as follows: Section 2.2 is assigned to reviewing the most important previous research done in the literature of the economics of immigration. In section 2.3 the data used for the study will be explained along with some descriptive statistics. Methodological approaches will be explained throughout section 2.4 and results will be discussed. Finally, section 2.5 will conclude.

2.2 Literature Review

Immigration has been a controversial topic among economists for a long time. The issue almost disappeared in the 1960's, but it became interesting among researchers again as immigrant inflows were significantly increased (Card, 2009). With all the refugees fleeing civil wars in the Middle East, the immigration related studies will be more than ever important.

In general, there are two main research streams which focus on the issues related to immigration effects. The first approach investigates the entire positive and negative effects of immigration on different components of the destination economy and labor market, such as wages, employment opportunities, job security, expenditures in social programs, etc. This approach sometimes also studies the changes in the labor market equilibrium and tries to find the new equilibrium after certain number of immigrants (with specific qualifications) enter the labor market of the destination country. This type of research is more common when the researcher studies immigrant friendly countries, such as the United States, Canada, Australia, etc. Borjas (1994, 1995b, 1999), Friedberg and Hunt (1995), Friedberg (2001), and Card (2001, 2009) are some important examples of studies carried out in this direction.

The other approach, which is also followed in this study, focuses on the earnings of immigrants in the host country's labor market. Many studies are carried out to check for any significant difference between earnings of immigrants and those of natives and also the economic assimilation of immigrants in the host countries while controlling for different socio-economic and demographic factors.

In this section, I mention some of the papers that made significant contributions to the literature of economics of immigration. I first start by those studies that employ cross-sectional approach, and then will review those using longitudinal (panel) approach. In order to cause a better understanding regarding the contribution of papers, I present papers of each group based on their chronological order.

Although so many people have worked on immigration related topics, Barry R. Chiswick (1978) is one of the pioneers in the field. Much of the research carried out afterwards is based on his methodology introduced in his 1978 paper. As explained earlier, the famous Mincerian earnings function is employed in his paper to examine the effect of foreign birth, length of time in the US (years since migration), and the US citizenship on the earnings of white men born outside of the United States. He uses the 1970 census cross-sectional data for his study. As explained in the paper, since in the 1970 data most of the foreign born are white, his analysis is restricted to whites to avoid a confounding of the effects of race and foreign origin on earnings. It is also limited to men due to the problem of estimating labor market experience for women. His results show that although immigrants initially earn less than natives, their earnings go up more rapidly with the US labor market experience, and after 10 to 15 years their earnings become equal, and then exceed, those of natives. Chiswick also concludes that earnings are unrelated to whether the foreign born immigrants are the US citizens or not. Long (1980) uses exactly the same dataset and the same model as Chiswick's (1978). The only difference is that he does not exclude females from his research. Since labor force participation of women may not be continuous over their life cycles and "age minus schooling minus 5" might be unreliable as a proxy for women's experience, in order to control for experience in the absence of work history, Long brings number of kids of a woman and her kids' age structures along with her marital status into the model. He concludes that the earnings of foreign born females are about 13% higher than their native counterparts.

Doubtful about Chiswick's (1978) results, George Borjas (1985) writes an influential paper to re-examine the empirical basis for two results found and reported in Chiswick's 1978 paper: the earnings of immigrants grow quickly as they assimilate into the United States; and this rapid growth results in immigrants overtaking natives in earnings within 10-15 years after immigration. Borjas (1985) uses both 1970 and 1980 US census datasets and again runs a Mincer-based wage equation on each of them separately. Instead of putting "years since migration" in the model, he uses dummy variables for different time periods of immigration, and by comparing the two estimates he shows that the cross-section regressions commonly used in the literature confound the actual assimilation impact with possible quality differences among immigrant cohorts. Instead of the fast growth found by the cross-section studies like that of Chiswick (1978), the cohort analysis predicts relatively slow rates of earnings growth for most immigrant groups. Borjas suggests the reason could be that the quality of those immigrants who migrated to the US from 1970 to 1980 was decreased.

Borjas (1994) in another well-cited paper studies three major issues in Economics of Immigration using United States 1970 and 1980 census data. First, how do immigrants perform in the host country's economy? Second, what impact do immigrants have on the employment opportunities of natives? Finally, which immigration policy is most beneficial to the host country? He checks for different immigrant cohorts and different generations and concludes: the relative skills of newer immigrants decreased over the post-war period; it is unlikely that new immigrants reach equality with the earnings of natives; immigration could be responsible for the decline in the earnings of unskilled natives during the 1980s; immigration policy matters since it can only let immigrants in who are more skilled, and less probably will participate in public assistance programs; and finally, there exists a strong correlation between the skills of immigrants and the skills of their descendants. So, he deduces that immigration has a long-lasting effect on the host country's economy. In another paper, Borjas (1995b) uses a simple economic framework to show how natives benefit from immigration. He argues that natives benefit from immigration mainly because of production complementarities between immigrant workers and other factors of production. Borjas concludes that these gains could be increased significantly if the United States follows an immigration policy that attracts more skilled immigrants.

In a very comprehensive study published in the handbook of labor economics, Borjas (1999) surveys the economic analysis of immigration by investigating the determinants of the immigration decision by workers in source countries and the impact of that decision on the host country's labor market and on immigrants' wages. In this paper, Borjas studies different aspects of immigration theoretically and empirically in a comprehensive way, and also reviews and critiques previous studies. He includes his own previous works and results on immigration as well. This paper could be used as a reference for further and deeper understanding of the economics of immigration.

Friedberg and Hunt (1995) review the empirical and theoretical studies done by different authors about different countries to investigate the impact of immigrants on host country wages, employment and growth. They conclude that despite the popular belief, the literature does not provide much evidence that immigrants have a large adverse impact on the wages and employment opportunities of the natives. In terms of growth, they conclude that theoretical literature on immigration and economic growth shows that the impact of immigrants on natives' income growth depends on the human capital levels of the immigrants, but empirical researches on this issue give conflicting answers.

Friedberg (2001) uses the natural experiment of 12% population increase in Israel between 1990 and 1994, and shows that despite the common belief, massive immigrations from the former Soviet Union did not have an adverse impact on native Israelis' earnings and employment opportunities.

Card (2001) studies the impacts of new immigration on the occupation-specific labor market using the 1990 US census data, and finds that immigrant inflows over the 1980's slightly reduced wages and employment rates of less skilled natives in traditional gateway cities. In another paper, Card (2005) focuses on the questions "does immigration reduce the labor market opportunities of less-skilled natives?" and "have immigrants who arrived after the 1965 Immigration Reform Act been successfully assimilated?" Card uses the 2000 US census data to answer the first question and March CPS data from 1995 to 2002 for the second one, and does not find enough evidence supporting that immigrants have a negative impact on less educated natives. On the question of assimilation he finds that few of immigrants, who come to the US without completed high school education, will ever catch up with the average earnings of natives. However, most of their US-born children will catch up with the children of natives. Card (2009) in another study presents an overview on the connection between immigration and wage inequality, focusing on comparing the US major cities. He comes up with the conclusion that within broad education classes, immigrant and native workers appear to be imperfect substitutes, with a large elasticity of substitution. So, immigration has slight effect on wage inequality among natives.

Chiswick and Miller (2011), interestingly, find "negative" assimilation among immigrants living in the United States if skills are highly transferable internationally. They use the traditional immigration assimilation model, and find that negative assimilation arises not from a deterioration of skills, but from a decline in the wages afforded by skills with longer period of residence. They use United State's census data from 1980, 1990, and 2000 to test the hypothesis on immigrants to the United States from English-speaking developed countries. They find that even after controlling for cohort quality effects, negative assimilation still occurs for immigrants in the sample. They also find similar results for immigrants from English-speaking developed countries living in Australia , and immigrants from Nordic countries in Sweden.

In one of his more recent studies, Borjas (2013) again examines the evolution of immigrant earnings in the US labor market using 1970-2010 census datasets. His results show that there are cohort effects both in the level of earnings, which means more recent cohorts generally have relatively lower entry wages, and also in the rate of growth of earnings, meaning that more recent cohorts have smaller rates of economic assimilation. He finds a part of this slowdown in wage convergence to be due to reduction in the rate of human capital accumulation especially English language proficiency in more recent cohorts. He claims that the English learning process is significantly slower for larger national origin groups, and the growth in the sizes of these groups accounts for about a quarter of the decline in the rates of human capital accumulation and economic assimilation.

Rodríguez-Planas and Vegas (2014) run a study for Spain using National Immigrant Survey (ENI-2007) to compare assimilation process of Moroccan immigrants with Ecuadorians and Romanians (the two other largest groups of immigrants in Spain). Employing Heckman-corrected estimates, their results show that after controlling for all socio-economic factors, Moroccans have higher wages at arrival and the difference does not decrease over time.

As explained earlier, in order to come up with more reliable and accurate assimilation rates cohort bias and survivor bias should be as much as possible avoided, and using longitudinal data is the only way for it. Since longitudinal datasets that provide researchers with enough information to do the analysis are rare, the number of studies carried out using panel data is significantly less compared to those using cross-section data. Some of the influential studies that use longitudinal approach are as follows:

Borjas (1989) analyzes the relationship between earnings and the extent of assimilation, cohort quality change, and out-migration experienced by the foreign-born population. He uses the longitudinal data available in the Survey of Natural and Social Scientists and Engineers from 1972 to 1978. His results show that there was a sizable decline in the skills of the population under study over the last two decades. His study also shows that return migration is more likely among immigrants who did not perform well in the United State's labor market compared to those who were successful.

Hu (2000) uses Health and Retirement Survey (HRS) longitudinal data source, which is a longitudinal survey of the population born between 1931 and 1941. He finds that the rate of increase of immigrants' earnings is overstated in census-based cross-sectional studies, and the gap between immigrant and native earnings for more recent arrival cohorts is larger than what was previously found.

Duleep and Dowhan (2002) use longitudinal data on earnings from a Social Security Administration (SSA) database matched to the 1994 March Current Population Survey. They study the initial earnings gap between natives and immigrants and also examine the trend over time in the foreign-born men's earnings growth. They find that for the year-of-immigration cohorts covered in their article (1960-1992) and for an important sub-set of the immigrant and native population, immigrant cohorts generally show higher earnings growth than do natives.

Chiswick et al. (2005b) develop a model of the occupational mobility of immigrants using data on males from the Longitudinal Survey of Immigrants to Australia. Their study shows that the initial occupational status of immigrants may be an unreliable approximation of their ultimate occupational achievement. Based on the results, immigrants with higher level of (transferable) skills, who are economic immigrants not refugees or family immigrants seem to have the most successful occupational adjustment through time. Moreover, they conclude living in an immigrant/ethnic concentration area seems to improve the occupational status of immigrants. In another paper, Chiswick et al. (2005a) study the determinants of the level and growth in earnings of adult male immigrants in their first 3.5 years in Australia using the same data. Using this data, they find that assimilation happens, and the cross-sectional approach provides a good estimate of the assimilations rate for immigrants which is derived from longitudinal analysis.

Beenstock et al. (2005, 2010) create a longitudinal dataset by matching immigrants in Israel's censuses for 1983 and 1995 and run multiple cross-sections, synthetic cohort methodology, and longitudinal analysis. They show that the results of the first two methods support the immigrant assimilation hypothesis (IAH), but the panel data analysis does not support it, and predicts that the earnings growth for immigrants varies inversely with duration.

Lubotsky (2007) uses a long matched panel data made through a joint project of the Social Security Administration, the Internal Revenue Service, and the Census Bureau. As a result, the 1990 and 1991 Survey of Income and Program Participation (SIPP) and the 1994 March Supplement to the Current Population Survey (CPS) got matched to annual earnings records from 1951 to 1997. He finds that the immigrant-native earnings gap closes by 10-15 percent during immigrants' first 20 years in the United States, which is about half as fast as typical estimates from repeated cross sections of the decennial census. His results indicate that immigration by low-wage immigrants has led past researchers to over-estimate the wage progress of immigrants who remain in the United States.

Banerjee (2009) examines the income growth of newly arrived immigrants in Canada using longitudinal data Survey of Labor and Income Dynamics (SLID) from 1999 to 2004. Her results indicate that recent immigrants face initial earnings disadvantage. However, while immigrants of European origins experience a period of catch-up early in their Canadian careers, which allows them to overcome this earnings disadvantage, "visible minority immigrants" (those immigrants who have visible differences in color, race, etc. compared to white natives) do not experience such a catch-up. She finds that this racial difference in recent immigrants' income growth is caused by the fact that visible minority immigrants receive lower returns to education, work experience and unionization. Furthermore, visible minority recent immigrants face greater penalties for speaking a non-official first language than do their white counterparts.

Izquierdo et al. (2009) use a longitudinal dataset "Continuous Sample of Working Histories" to examine the earnings assimilation of immigrants in Spain. They show that immigrants reduce the wage gap compared to natives by 15% during the first 5-6 years after arrival, but the earnings gap will not be gone completely. Based on their results, assimilation is faster for South American and European immigrants compared to Africans.

Hall et al. (2010) using the 1996-1999 and 2001-2003 panels of the Survey of Income and Program Participation (SIPP) estimate wage differences for four groups: documented Mexican immigrants, undocumented Mexican immigrants, US-born Mexican Americans and native non-Latino whites. Their results show that after controlling for other factors, there exist 8% and 4% wage differences between documented and undocumented Mexican immigrant men and women, respectively. They also find large differences in returns to human capital, with undocumented Mexican immigrants having the lowest wage returns to human capital and having very slow wage growth over time.

Abramitzky et al. (2012) study the assimilation of European immigrants in the United States labor market during the so-called "Age of Mass Migration" (1850-1913) using a newly-assembled panel data, and show that the average immigrant did not experience a substantial earnings difference upon first arrival. Their results also show that immigrants experienced occupational advancement at the same rate as natives. Moreover, they show that assimilation patterns vary across sending countries and persist in the second generation.

Picot and Piraino (2013) use longitudinal tax data which is matched to immigrant landing records to study the effect of selective attrition on the estimated earnings assimilation of immigrants to Canada. Contrary to findings in the existing literature, they find that the immigrant-native earnings gap closes at the same pace in longitudinal and cross-sectional data. They also find that low-earning immigrants are likely to leave the cross-sectional samples over time, but the same is true for the native born. Their study shows that labor market participation patterns of immigrants is similar to those of native Canadians.

Lemos (2013) uses the large and long Lifetime Labor Market Database (LLMDB) of the United Kingdom and estimates the immigrant-native earnings gap at entry and over time for the UK between 1978 and 2006 with both cross-sectional and longitudinal approaches. She separately estimates cohort and assimilation effects, and also estimates the associated immigrants' earnings growth rate and immigrant-native earnings convergence rate. Her results suggest that immigrants from more recent cohorts do better than earlier ones at entry and their earnings also catch up faster with natives' earnings. Gagliardi and Lemos (2015) also use the same data from 1981 to 2006 to investigate the evolution of the immigrant-native earnings gap across the entire earnings distribution, across cohorts and across nationalities. They control for both cohort-specific effects and nationality-specific effects. Their results show little evidence of large or persistent earnings inequalities across cohorts or across nationalities. By that they conclude that recipient labor markets primarily reward individuals' characteristics regardless of their immigration status. However, when they investigate the change of the immigrant-native earnings gap over time, they find that immigrants from different continents and cohorts have very different assimilation patterns.

Kaushal et al. (2015) use Canadian Survey of Labor and Income Dynamics (SLID) and the US Survey of Income and Program Participation (SIPP) from 1996 to 2008 and study the short-term routes of employment, hours worked, and real wages of immigrants in Canada and the United States. Their models with personal fixed effects show that, on average, immigrant men in Canada do not experience any relative growth in these three outcomes compared to men born in Canada. Immigrant men in the US, on the other hand, experience positive annual growth in all three domains relative to US-born men. They also compare longitudinal and cross-sectional results and find that the latter over-estimate wage growth of earlier arrivals, presumably, reflecting selective return migration.

2.3 Data and Descriptive Statistics

As mentioned earlier, the 2003, 2006, and 2008 cycles of the under-explored National Survey of College Graduates (NSCG), which is collected by the US Census Bureau (under the auspices of the National Science Foundation (NSF)), is used in this study. The National Survey of College Graduates is a longitudinal biennial survey that provides data on the nation's college graduates, with particular focus on those in the science and engineering fields. The program has been conducted since the 1970's. The survey samples are individuals living in the United States during the survey reference week, have at least a bachelor's degree, and are under the age of 76. The 2003 and 2010 cycles of the NSCG provided coverage of the nation's college-educated population as of the survey reference date. In addition to the 2003 and 2010 survey cycles, the NSCG was conducted biennially or triennially in the period 2000-2009. For the within-decade iterations of the NSCG (2006 and 2008), the survey focused on the science and engineering (S&E) workforce component of the college-educated population. The 2003 NSCG selected its sample from the 2000 decennial census long form respondents who indicated they had a bachelor's degree or higher in any field of study. The 2003 NSCG survey respondents served as the sample source for future survey cycles within the 2000 decade (i.e. the 2006 and 2008 cycles). Only those who were recipients of a bachelor's degree or higher in a science, engineering, health-related, or S&E-related field prior to April 2000 or were employed in a science, engineering, health-related, or S&E-related occupation as of October 2003 were followed in 2006 and 2008 cycles.

NSCG is a unique source for examining various characteristics of college-educated individuals, including occupation, salary, the last five highest degrees and their majors, whether each degree was received in the US, the type of entry visa for immigrants and their current residency status, and detailed demographic information. The 2010 wave of NSCG selected a part of its sample from the 2009 American Community Survey (ACS) respondents who indicated that they had a bachelor's degree or higher in any field of study. The remaining portion of the 2010 NSCG sample was selected from respondents to the 2008 NSCG[2].

Although it costed me one less point in time in the panel I used for my longitudinal analysis, for two reasons I decided not to include the 2010 cycle in my panel: 1 - the attrition rate from individuals present in the 2008 cycle to the 2010 cycle was more than 50 percent, and 2 - the recent financial/economical crisis which could have changed the out-migration pattern of the US immigrants just started to happen after the 2008 cycle and could have affected the values of the 2010 cycle.

Most of administrative datasets provide researchers with information about only few characteristics of each observation, and in most of the cases educational achievement is not a part of them. The NSCG provides the researcher with much in-detail information for each individual. The amount of detailed information given by NSCG data about the education and employment of each individual in the sample is exceptional. Its focus on college-educated people with science and engineering (S&E) fields is also an advantage of this dataset. The only drawback of this longitudinal (panel) data is the few points of time available for each individual. Nevertheless, knowing the fact that there are not many longitudinal datasets available for such studies, NSCG could be considered as a valuable source. Moreover, some studies such as Beenstock et al. (2005, 2010) have done their analyses with even fewer time points.

The 2003 survey cycle in total includes 100,402 individuals out of which 76,778 either have education in science and engineering (S&E) majors or work in such fields. The 2006 and 2008 cycles include 51,694 and 45,033 observations, respectively. Each of these individuals either hold degrees in S&E or work in S&E fields. For this study, I only keep those S&E-related individuals who are living and working inside the United States during survey reference weeks and are 65 years old or younger. To get more precise results in my study, hourly wage of each individual is calculated using the reported annual salary, number of weeks worked per year, and number of hours worked per week. It seems that some respondents have confused annual weeks and months, or weekly and daily hours or for any other reason have reported too big or too small annual salaries. Based on the 2013 report of the Labor Department of

[2]Ref.: http://www.nsf.gov/statistics/srvygrads/#sd

the United States, less than about 2% of annual wages in 2013 were more than 200,000 USD, while in the three cycles more than 3% of people have reported bigger annual wages and about the same percentage of respondents have reported zero or very low incomes. Anyway, it should be considered that the wages reported in the NSCG cycles are earnings of highly educated individuals. So, I do another level of filtration and only keep those individuals who earn more than or equal to the federal minimum wage of year 2003 ($5.15) and less than $250 per hour[3]. For instance, . Dropping observations that do not meet the criteria explained above and also removing individuals with some conflict or error in their responses from the data, leaves me with 55,929 observations in the 2003 cycle, 38,881 in 2006, and 31,634 observations in the 2008 cycle. Out of these observations, the 2003 cycle has 35,505 men and 20,424 women; 2006 cycle includes 24,976 men and 13,905 women; and 2008 cycle has 20,462 male observations and 11,172 females. In the 2003 cycle: 76.18% of observations are natives; 15.16% are naturalized US citizens; 6.13% are permanent residents (Green Card holders); and the remaining 2.53% are temporary residents of the US (US visa holders). The respective percentages for the 2006 cycle are 76.58, 16.26, 6.03, and 1.13 percent, and for the 2008 cycle: 76.39, 18.62, 4.56, and 0.43 percent.

A common problem in longitudinal (panel) datasets that could affect the results of longitudinal studies (especially immigration studies) is the attrition problem. It should be noted that even a high attrition rate might cause no bias in the estimates if it leaves the researcher with adequate observations for the analysis and also happens randomly. However, existence of some specific patterns or kind of selectivity in causing attritions can be troublesome. Attrition problem can be even more complicated and problematic when the longitudinal data includes a significant number of observations who are immigrants. Various factors can cause attrition in datasets and most of them are common between immigrants and natives, but there is one which is immigrant-specific and could cause attrition among immigrant individuals of the sample: immigrants might out migrate from the destination country selectively and either go back home or migrate to other countries. Obviously, depending on whether the more able and more successful immigrants are moving out or less able ones, estimated coefficients could be negatively or positively biased. That is why it is very important that the researcher knows about the source of attrition. Unfortunately, the NSCG does not provide

[3]Altogether, 707 observations (474 natives and 233 immigrants) out of 127,504 observations of the three cycles together, have wages higher than $250 per hour (about 0.55% of the whole population, 0.49% of natives and 0.77% of immigrants). I ran sensitivity analysis to see how different my regression results would be without removing observations with hourly wages higher than $250, and did not find considerable changes in estimated coefficients.

researchers with any information regarding the attrition sources[4].

In the data that I am using, out of 55,929 observations in the cycle 2003, only 30,918 of them are followed all the way to 2008[5]. From this number, 24,448 observations belong to US-born and 6,470 observations relate to immigrants. The interesting fact about immigrants, however, is that while the 2003 cycle has 4,156 naturalized citizens, 1,730 permanent residents, and 584 temporary residents (visa holders), these numbers change to 4,487, 1,731, and 252 in 2006 and then to 5,084, 1,265, and 121 in 2008, respectively. The reason is obvious though. With more time resided in the United States, some of those who held visa in 2003 change status to permanent resident in 2006 or 2008 and some of those who were permanent residents at the time, become naturalized US citizens[6]. From the whole immigrant population of the data who were present in all of the three cycles (6,470), 2,352 of them migrated to the United States on a permanent residence visa, 1,081 of them came on a work visa (such as H-1B), 1,972 on student visa, 612 on a dependent visa, and the remaining 453 on other types of visa.

Tables 2.1, 2.2, and 2.3 depict some descriptive statistics on some of the important variables used in the analysis using the weighted sample (and sub-samples). As could be seen in the three tables, immigrants on average have a larger hourly wage compared to natives, while the average ages of immigrants and natives are similar with just a little bit of margin for natives. The kernel density plots in graphs 1 through 6 of figure 2.1 show that immigrants' and natives' wages are distributed similarly, so there will be no need to extend the wage analysis beyond mean regressions in this study.

According to the left sides of tables 2.1, 2.2, and 2.3, amongst immigrants, naturalized citizens have the highest hourly wages which could be because of the following four reasons: 1) their mean of age is higher; 2) on average, they have more years since migration (YSM); 3) they migrated to the US when they were younger so they were exposed to American

[4] I contacted the NSCG project officer at the National Science Foundation several times in order to get more information (if any) about the attrition in the data, but I never got replied.

[5] Since individuals in this data are highly educated, the opportunity cost of filling the NSCG's long and time taking survey and sending it back to the NSF is quite high. So, seeing an attrition rate higher than other longitudinal datasets should not be surprising. As can be seen in table 2.1, the attrition rate for natives and immigrants is higher among those with higher degrees (holders of doctorate degrees for instance have a very high attrition rate). It is also interesting to notice that the attrition rate for both natives and immigrants is higher among those with higher hourly wages.

[6] 349 individuals change status to permanent resident from 2003 to 2006; 137 individuals change status to permanent resident from 2006 to 2008; 394 individuals change status from permanent resident to citizen from 2003 to 2006; and 610 individuals change status from permanent resident to citizen from 2006 to 2008.

education/experience/culture from a younger age; and 4) they have the citizenship which might be advantageous in the US labor market. The second highest hourly wage (which is not much different from that of naturalized citizen immigrants) is for permanent residents of the United States, and the lowest amongst immigrants is earned by temporary residents which are on average the youngest group with the least years since migration.

Looking at the number of years since migration to the United States for different groups of immigrants based on their residency status, a logical pattern could be seen. Naturalized citizens on average have the highest number of YSM; permanent residents' (weighted) mean for years since migration is at the second level; and temporary residents of the US who hold visas have been living in the United States for the shortest period.

As can be seen in tables 2.1, 2.2, and 2.3, immigrants on average have significantly higher education levels compared to natives. Both natives and immigrants mostly hold bachelor's degrees in all the three cycles (except for the 2008 cycle in which number of immigrants with a master's degree is higher than the number of immigrants holding a bachelor's degree as their highest level of education), but the ratio of undergraduate and postgraduate degrees for immigrants and natives are significantly different, showing higher education in immigrants. In the 2003 cycle, around 61% of all immigrants have earned their highest degrees from a university inside the United States. 70% of naturalized citizens hold a US earned highest degree, while only about 46% of permanent residents and 44% of temporary residents have earned their highest degrees from a school inside the US. In the cycle 2008, the percentages are declined to 57% for all immigrants; 63% for naturalized citizens; 34% for green card holders; and about 40% for temporary residents[7].

Men form about 62% of the us-born observations in the 2003 sample, while about 67% of the entire immigrant observations are male. Regarding sub-groups of immigrants, share of men gets to 65%, 70%, and 76% for naturalized citizens, permanent residents, and temporary residents, respectively. Over next cycles, the share of male observations from the entire sample gets even more unbalanced. For instance, in 2008 cycle we see 64% men among natives, and 64%, 75%, and 82% men among citizen immigrants, Green Card holders, and visa holders, respectively. This noticeable difference between natives and immigrants with respect to share of genders could be a sign of the possibility that for various reasons immigration (which is per se a tough process) is more done by men than women.

[7]Percentages are calculated based on all of the observations, rather than just those present in all three cycles.

TABLE 2.1: Summary Statistics (weighted) for the 2003 Cycle

	All observations						**Those who are present in all cycles**					
	Whole Sample	**Native**	**Immigrant**	Naturalized Citizen	Permanent Resident	Temporary Visa	**Whole Sample**	**Native**	**Immigrant**	Naturalized Citizen	Permanent Resident	Temporary Visa
Number of observations	55,929	42,606	13,297	8,480	3,427	1,416	30,918	24,448	6,470	4,156	1,730	584
Hourly Wage ($)	33.27 (18.12)	32.66 (17.86)	35.21 (18.79)	36.42 (19.17)	34.22 (17.94)	30.41 (17.54)	32.22 (20.60)	31.94 (20.32)	34.01 (22.22)	34.15 (20.88)	34.47 (25.88)	31.42 (19.84)
Age	44.14 (9.79)	44.25 (9.85)	43.81 (9.59)	46.27 (9.46)	41.06 (8.21)	35.82 (6.93)	43.13 (8.89)	43.22 (8.92)	42.60 (8.67)	44.39 (8.56)	40.31 (7.77)	35.48 (6.68)
Highest Level of Education*												
Bachelor's	49.40	53.50	35.62	37.96	33.32	26.48	55	58.90	40.25	43.09	35.95	32.71
Master's	31.30	30.38	34.53	33.31	35.95	38.91	32.98	31.34	39.18	38.40	41.45	38.01
Doctorate	10.95	8.07	20.77	17.89	25.24	28.11	2.97	1.09	10.08	6.45	15.78	19.01
Professional	8.35	8.05	9.08	10.84	6.27	6.43	9.05	8.66	10.49	12.05	6.82	10.27
YSM	-	-	19.73 (11.64)	24.42 (10.82)	13.29 (8.41)	7.54 (4.82)	-	-	19.05 (11.06)	23.09 (10.24)	12.60 (8.43)	7.03 (4.29)
Age at Arrival	-	-	23.74 (10.54)	21.47 (11.01)	27.62 (8.91)	27.85 (6.39)	-	-	23.56 (10.46)	21.30 (10.75)	27.72 (8.80)	28.45 (6.41)
Males*	63	61.70	67.30	64.42	70.12	76.48	64.59	63.71	67.93	64.41	72.43	79.62
Married*	74.83	72.82	81.44	81.37	85.29	71.05	80.23	79.08	84.59	84.29	87.75	77.40

YSM stands for "Years since Migration"; Numbers in parentheses are standard errors; Numbers of observations reported are not weighted.
*Numbers are in percentages.

TABLE 2.2: Summary Statistics (weighted) for the 2006 Cycle

	All observations						Those who are present in all cycles					
	Whole Sample	**Native**	**Immigrant**	Naturalized Citizen	Permanent Resident	Temporary Visa	**Whole Sample**	**Native**	**Immigrant**	Naturalized Citizen	Permanent Resident	Temporary Visa
Number of observations	38,881	29,774	9,107	6,323	2,343	441	30,918	24,448	6,470	4,487	1,731	252
Hourly Wage ($)	36.86 (25.59)	36.61 (25.40)	38.08 (26.50)	38.77 (27.39)	37.16 (24.63)	33.37 (22.23)	35.34 (19.60)	35.09 (19.55)	36.98 (19.85)	37.23 (20.27)	36.86 (18.81)	33.41 (18.37)
Age	45.52 (9.34)	45.68 (9.38)	44.74 (9.08)	46.40 (9.02)	41.77 (7.96)	37.02 (7.31)	46.13 (8.89)	46.22 (8.92)	45.60 (8.67)	47.11 (8.58)	42.56 (7.69)	37.90 (7.06)
Highest Level of Education*												
Bachelor's	52.25	56.18	38.96	41.44	34.78	25.62	53.98	57.79	39.61	41.85	35.24	29.76
Master's	34.22	32.58	39.31	38.79	41.40	35.60	33.73	32.18	39.60	39.36	40.73	36.11
Doctorate	4.12	2.07	10.78	7.20	16.99	29.25	3.13	1.25	10.23	7.06	16.87	21.03
Professional	9.41	9.17	10.95	12.57	6.83	9.52	9.15	8.78	10.56	11.72	7.16	13.10
YSM	-	-	21.90 (11.13)	25.33 (10.53)	14.81 (8.42)	10.31 (4.55)	-	-	22.05 (11.06)	25.31 (10.44)	14.72 (8.37)	10.04 (4.36)
Age at Arrival	-	-	20.97 (11.91)	19.42 (11.97)	24.49 (11.17)	24.47 (10.01)	-	-	23.56 (10.46)	21.80 (10.78)	27.84 (8.37)	27.86 (6.98)
Males*	64.37	63.46	66.76	63.62	73.24	77.32	64.59	63.71	67.93	64.25	75.33	82.54
Married*	80.89	79.48	85.48	85.61	86.90	76.19	80.23	79.08	84.59	86.38	87.46	76.59

YSM stands for "Years since Migration"; Numbers in parentheses are standard errors; Numbers of observations reported are not weighted; Weighted averages of the right side are calculated with the weights given in the 2003 cycle.

* Numbers are in percentages.

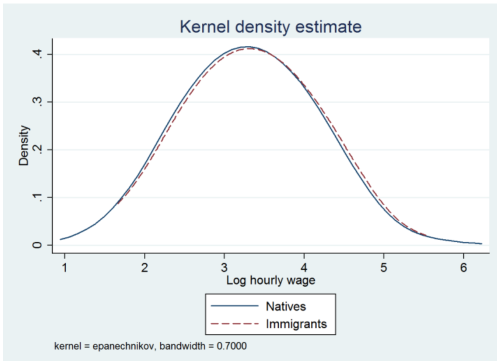
TABLE 2.3: Summary Statistics (weighted) for the 2008 Cycle

	All observations						Those who are present in all cycles					
	Whole Sample	**Native**	**Immigrant**	Naturalized Citizen	Permanent Resident	Temporary Visa	**Whole Sample**	**Native**	**Immigrant**	Naturalized Citizen	Permanent Resident	Temporary Visa
Number of observations	31,634	24,166	7,468	5,889	1,442	137	30,918	24,448	6,470	5,084	1,265	121
Hourly Wage ($)	40.96 (26.39)	40.50 (25.78)	43.17 (29.06)	43.32 (29.06)	42.88 (29.77)	40.20 (21.39)	39.17 (21.11)	38.75 (20.89)	41.83 (22.27)	41.87 (22.24)	42.17 (22.78)	37.88 (18.46)
Age	46.79 (8.89)	46.94 (8.93)	46.04 (8.63)	46.84 (8.63)	43.65 (7.92)	38.92 (7.35)	48.13 (8.89)	48.22 (8.92)	47.60 (8.67)	48.48 (8.62)	44.79 (7.88)	39.22 (7.28)
Highest Level of Education*												
Bachelor's	51.66	55.76	38.38	39.77	33.22	32.85	52.65	56.44	38.33	39.65	33.44	33.88
Master's	34.79	33.05	40.44	40.79	39.18	38.69	34.69	33.14	40.54	41.37	37.47	38.02
Doctorate	3.51	1.51	10	7.57	19.14	18.25	3.43	1.56	10.49	7.87	20.32	18.18
Professional	10.04	9.68	11.18	11.87	8.46	10.22	9.23	8.86	10.63	11.11	8.77	9.92
YSM	-	-	23.95 (11.01)	25.96 (10.72)	16.97 (8.85)	12.51 (5.15)	-	-	24.05 (11.06)	26.03 (10.77)	16.85 (8.77)	12.33 (4.94)
Age at Arrival	-	-	19.27 (12.36)	18.39 (12.22)	22.34 (12.57)	23.30 (10.23)	-	-	23.56 (10.46)	22.45 (10.67)	27.94 (8.50)	26.89 (6.64)
Males*	64.68	64.04	66.75	64.37	74.97	82.48	64.59	63.71	67.93	65.18	77.71	80.99
Married*	82.64	81.40	86.65	86.72	87.52	74.45	80.23	79.08	84.59	87.02	87.35	75.21

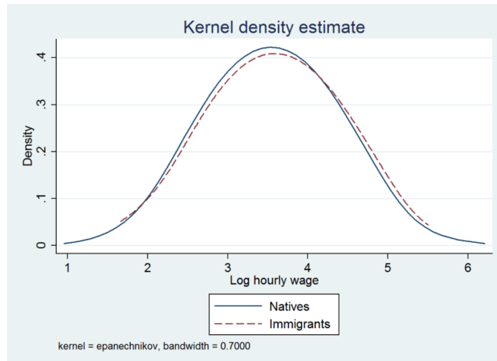
YSM stands for "Years since Migration"; Numbers in parentheses are standard errors; Numbers of observations reported are not weighted. Weighted averages of the right side are calculated with the weights given in the 2003 cycle.
* Numbers are in percentages.

Figure 2.1: Kernel density distributions

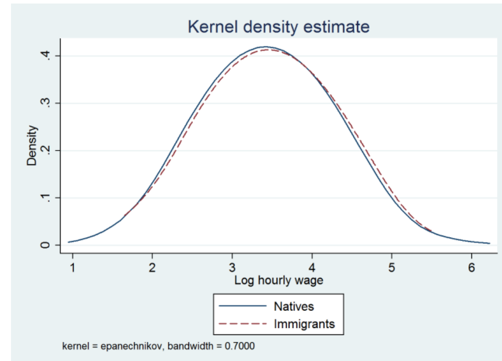
Graph1. Kernel density distributions with bandwidth of 0.7 for the 2003 cycle (all observations)



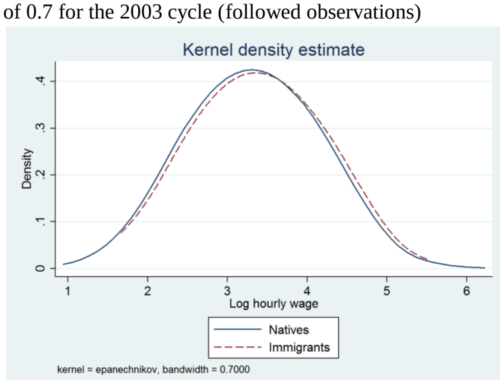
Graph2. Kernel density distributions with bandwidth of 0.7 for the 2006 cycle (all observations)



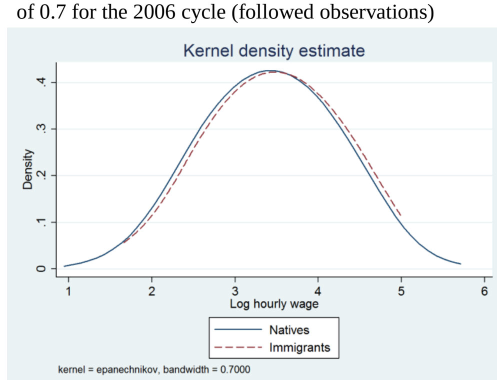
Graph3. Kernel density distributions with bandwidth of 0.7 for the 2008 cycle (all observations)



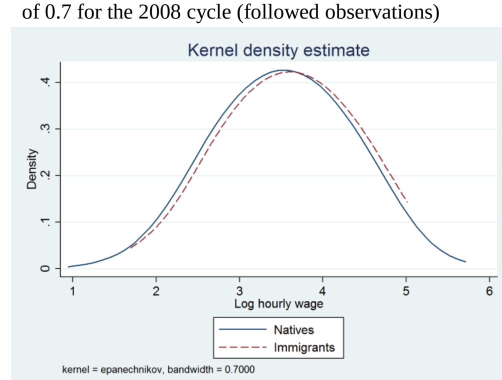
Graph4. Kernel density distributions with bandwidth of 0.7 for the 2003 cycle (followed observations)



Graph5. Kernel density distributions with bandwidth of 0.7 for the 2006 cycle (followed observations)



Graph6. Kernel density distributions with bandwidth of 0.7 for the 2008 cycle (followed observations)



Ratio of people with physical disability to the whole sample is another difference between natives (about 6%) and immigrants (about 4.5%) that simply could be because people with disabilities do not find the immigration as much doable and/or rewarding as healthy individuals find it to be.

2.4 Methodology and Results

2.4.1 Cross-sectional Multivariate Analysis

As described in the literature (Chiswick (1978), Borjas (1999), etc.), immigrants and natives have different human-capital earnings functions:

$$\log w_n = \beta_0 + \boldsymbol{X_n \beta_1} + \beta_2 A_n + \beta_3 A_n^2 + \varepsilon_n \tag{2.1}$$

$$\log w_i = \alpha_0 + \boldsymbol{X_i \alpha_1} + \alpha_2 A_i + \alpha_3 A_i^2 + \alpha_4 YSM_i + \alpha_5 YSM_i^2 + \phi_i \tag{2.2}$$

where w is wage (could be annual, monthly, weekly, or hourly), $\boldsymbol{X}$ is a vector of socio-economic and demographic characteristics (including education, gender, marital status, region of employment, etc.)[8], A gives the age, and YSM counts number of years that the immigrant has resided in the host country. Index n is used for natives and index i is used for immigrants. For immigrants, based on Borjas (1985) discussion, the arrival cohort should be also controlled for in equation 2.2 to capture the differences in the skills and characteristics of different groups of immigrants who entered the host country over different time periods[9]. Since equation 2.2 controls for age and age-squared of immigrants, coefficients α_4 and α_5 measure the differential value that the host country's labor market attaches to the time spent in the host country in contrast to the time spent in the home country (Borjas, 1999). It

[8] According to Borjas (1999), one of the reasons of disagreement in the empirical literature over the relative economic status of immigrants in the United States is that different studies use different controlling variables. As a result, the base group differs drastically from study to study. For example, many studies include a worker's education (as years of schooling or level of the highest degree) in the vector X, and consequently, different effects are measured relative to native workers who have the same schooling. As Borjas discusses, a part of the adaptation process experienced by immigrants might include the acquisition of additional schooling. By controlling for schooling observed at the time of the survey, the analysis might hide the fact that there might be a big wage convergence between immigrants and natives (Borjas, 1999). In this study, since the individuals presented by the data are highly educated, it is very important to standardize the education level and compare immigrants and natives with the same level/quality of education.

[9] I divided immigrants into 7 separate groups based on their arrival year: 1-those who entered the US before 1949; 2-between 1950 and 1959; 3-between 1960 and 1969; 4-between 1970 and 1979; 5- between 1980 and 1989; 6- between 1990 and 1999; and 7- those who arrived after year 2000.

is crucial to let immigrants and natives have different coefficients for age and age-squared because of the notion of specific human capital. "After all, it is very unlikely that a year of pre-migration 'experience' for immigrants has the same value in the host country's labor market as a year of experience for the native population" (Borjas, 1999).

According to Borjas (1999), the rate of wage convergence/divergence between immigrants and natives is defined as follows:

$$(2.3) \quad CR = \frac{\partial \log w_i}{\partial time} - \frac{\partial \log w_n}{\partial time} = \alpha_2 + 2\alpha_3 \bar{A}_i + \alpha_4 + 2\alpha_5 YSM_i - \beta_2 - 2\beta_3 \bar{A}_n$$

where $\bar{A}_i$ and $\bar{A}_n$ are average ages of immigrants and natives, respectively.

I use the following least squares model with robust standard errors for my cross-sectional analysis to find the earnings difference between natives and different groups of immigrants at entry, and also to calculate the economic convergence/divergence rate between them:

$$(2.4) \quad \log\, w_j = \gamma_0 + \boldsymbol{I_j \gamma_1} + \boldsymbol{X_j \gamma_2} + \gamma_3 A_j + \gamma_4 A_j^2 + \gamma_5 YSM_j + \gamma_6 YSM_j^2 + f_t + \delta_j$$

where $\boldsymbol{I}$ is a matrix containing three dummy variables for the three different groups of immigrants (based on the current residency status) present in the data: naturalized citizens, permanent residents, and temporary residents. Obviously, one and only one of these dummies will take one for each immigrant individual in the sample under study and they will all take zeros when it comes to a native. Hence, the estimated coefficient on the indicator (dummy) variable of any above mentioned immigrant status (if significant) represents the log earnings difference between US-born (native) workers and that group of immigrants.

As mentioned earlier, since the NSCG data has provided me with the number of hours worked per week and number of weeks worked per year for each observation, I calculated the hourly wages from the annual salaries reported. So, the dependent variable of the model is the log of hourly wage. $\boldsymbol{X}$ is a matrix which includes the socio-economic and demographic attributes of individuals in the sample such as the highest university degree (bachelor's degree, master's degree, doctorate degree, or professional degree), whether the highest degree was earned in the United States, field of study of the highest degree, arrival cohort, sex, race, birth place, whether from an English speaking country[10], marital status, whether individual has children, employment sector, employment (job) location/region, employer size, full-time/part-time, and physical disability indicator. YSM, as explained earlier, is the number of years since

[10] I made this dummy variable based on the country of origin of immigrants. The dummy variable takes 1 if the immigrant is born in an English speaking country and zero otherwise. Obviously, it takes one for natives.

migration to the United States and takes zero for natives.

"Friedberg (1992) argued that the generic model ignores an important aspect of immigrant wage determination: the role of age-at-arrival in the host country. The US data suggest a strong negative correlation between age-at-arrival and entry earnings. The identification problem, however, does not disappear when the entry wage of immigrants depends on age-at-migration. Rather, it becomes more severe"(Borjas, 1999). It is obvious that perfect collinearity exists between age, years since migration, and age at arrival, or among age at arrival, years since migration, and year of observation. In order to bring age at arrival into the model (besides arrival cohort), one need to impose an additional restriction. One possible restriction is to assume that the coefficients of "age" and "age-squared" variables are the same for immigrants and natives. The problem, though, is that such an assumption is very restrictive, and contradicts the notion of specific human capital. As mentioned earlier, it is very improbable that a year of pre-migration experience for immigrants has the same value in the host country's labor market as a year of experience for natives (Borjas, 1999). An alternative approach to address such an identification problem is modeling the age at arrival effect as a step function (using a categorical variable): persons who migrate as children, middle aged individuals, or old people face different opportunities in the host country than those who migrate as adults (Borjas, 1999). It is common between researchers to use such a categorical variable to break the collinearity, but it does not solve the real problem[11].

In order to come up with more precise results and take advantage of all of the observations in 2003, 2006, and 2008 cycles, I pooled observations of the three cycles[12] and using the consumer price index (CPI) reported by the Bureau of Labor Statistics (BLS), recalculated the 2006 and 2008 hourly wages in 2003 prices[13]. So, equation 2.4 is run as a pooled OLS. In order to control for and capture the differences in year-specific economic effects (like national and international macroeconomic shocks) on earnings between years 2003, 2006, and 2008, f_t is used in the model as time fixed effect[14].

[11] I ran my model with such a categorical variable as well. Not only were the estimated coefficients for the age-at-arrival insignificant in the main model and in all of auxiliary regressions, but also, including it did not change the other estimated coefficients considerably.

[12] All of individuals in all the three cycles who are in S&E fields, are living and working inside the United States during survey reference weeks, and are 65 years old or younger.

[13] According to the BLS (http://www.bls.gov/cpi/cpi_dr.htm), CPI in 2003, 2006, and 2008 are 184.0, 201.6, and 215.30, respectively.

[14] To solve the identification problem caused by collinearity among calendar year of the cycle, arrival year cohort, and years since migration, I keep the period effect (f_t) the same for both immigrants and natives (Borjas, 1994).

Table 2.4 depicts the results of running equation 2.1 on the native observations of the data and equation 2.2 on immigrant observations (whole group and different sub-groups based on their first entry visa type), separately. As can be seen, the estimated coefficients for age and age-squared for natives and immigrants are considerably different. So, as Borjas (1999) emphasized, forcing the same age-earnings profile to both natives and immigrants by assuming that both groups have the same coefficients for age and age-squared is a very strong and also misleading restriction. Hence, I run equation 2.4 on all of the observations in the pooled data without such a restriction and let natives and immigrants have different age-earnings profiles. In order to let that happen, interaction terms are employed[15].

This unrestricted model is then run four more times with four different sub-set of observations to compare natives with the four sub-groups of immigrants (based on the first entry visa type): The first sub-set of the pooled data includes all native observations plus only those immigrants who migrated to the United States using a permanent residence (immigrant) visa; The second one has natives and those who moved to the US for the very first time on a work visa; The third one includes natives and those immigrants with student visa as their first entry visa; and finally, the fourth sub-set of data has natives and those who migrated to the United States on a dependent visa.

2.4.2 Cross-section results

Table 2.5 reports the results from running equation 2.4 on the whole data and its sub-sets. Column (1) shows the results using the whole pooled sample. Columns (2) and (3) show results for men and women, respectively, and columns (4) through (7) demonstrate the cross-section results after running the model on the four above mentioned sub-sets of data: natives and immigrants who migrated on immigrant visa, natives and immigrants who migrated on work visa, natives and immigrants who migrated on study visa, and natives and immigrants who migrated on dependent visa.

Before getting to the results reported in table 2.5, it is worthwhile to talk a bit about one of the important findings from running equation 2.4, which is in contrast with what similar studies have found. Interestingly, the estimated coefficients for the dummies made for cohorts

[15]Two dummy variables, Imm and Nat were made. The former takes one for immigrants and zero for natives, and the latter takes one for natives and zero for immigrants. Then A and A^2 in equation 2.4 were replaced by the interactions of Imm and Nat with A and A^2.

Table 2.4: Separate-run cross-section results for natives and different groups of immigrants based on their first visa type

	Natives	**All immigrants**	**PR visa**[1]	**Work visa**[2]	**Study visa**[3]	**Dependent visa**[4]
Current Immigration Status						
Permanent Resident	-	0.004	-0.037*	-0.003	-0.049***	-0.004
		(0.008)	(0.017)	(0.017)	(0.012)	(0.027)
Temporary Resident	-	-0.054***	-	-0.033	-0.193***	-0.108*
		(0.014)		(0.025)	(0.020)	(0.053)
Age	0.053***	0.020***	0.026***	0.022**	0.004	0.027**
	(0.002)	(0.003)	(0.005)	(0.008)	(0.006)	(0.011)
Age Squared	-0.00051***	-0.00022***	-0.00029***	-0.00024**	-0.00007	-0.00028*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
YSM	-	0.011***	0.021***	0.003	0.015***	0.004
		(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
YSM Squared	-	-0.00008***	-0.00022***	0.00000	-0.00010*	-0.00000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Adj. R-Squared	0.225	0.214	0.216	0.179	0.266	0.278
No. of Observations	96546	28269	9957	4244	9473	2644

* $p<0.05$, ** $p<0.01$, *** $p<0.001$; Numbers in parentheses are robust standard errors; 1.Model is run only on those observations who first migrated to the US on an immigrant (permanent residence) visa; 2.Model is run only on those observations who first migrated to the US on a work visa; 3.Model is run only on those observations who first migrated to the US on a study visa; 4.Model is run only on those observations who first migrated to the US on a dependent visa; Numbers in parentheses are robust standard errors; Dependent variable is the log of adjusted hourly wage (2003 prices); Other independent variables that exist in all regressions but are not reported are as follows: Highest level of education(Bachelor's degree(baseline), Master's degree, Doctorate degree, and Professional degree), whether the highest degree is earned in the US, field of education, gender, marital status, whether has child(ren), whether English is the mother tongue, physical disability, place of birth, employment region, employer size (in terms of number of employees), employment sector, and survey year fixed effect.

Table 2.5: Cross-section Results using an unrestricted model on pooled data

	All	**Men**	**Women**	**PR**	**Work**	**Study**	**Dependent**
Immigration Status							
Naturalized Citizen	0.832***	1.001***	0.565***	0.479***	1.190***	1.244***	0.471*
	(0.074)	(0.091)	(0.127)	(0.110)	(0.186)	(0.132)	(0.211)
Permanent Resident	0.841***	1.022***	0.542***	0.447***	1.197***	1.203***	0.458*
	(0.073)	(0.091)	(0.126)	(0.110)	(0.184)	(0.130)	(0.208)
Temporary Resident	0.775***	0.940***	0.523***	-	1.165***	1.057***	0.332
	(0.072)	(0.089)	(0.125)		(0.181)	(0.127)	(0.209)
Age (Natives)	0.054***	0.058***	0.053***	0.053***	0.054***	0.054***	0.053***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Age (Immigrants)	0.020***	0.018***	0.028***	0.027***	0.014	0.001	0.032***
	(0.000)	(0.000)	(0.000)	(0.005)	(0.008)	(0.006)	(0.009)
Age Squared (Natives)	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age Squared (Immigrants)	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
YSM	0.010***	0.009***	0.012***	0.020***	0.002	0.013***	0.006*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
YSM Squared	-0.000***	-0.000*	-0.000***	-0.000***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Adj. R-Squared	0.224	0.213	0.193	0.224	0.226	0.230	0.226
Number of Observations	124815	80004	44811	106503	100790	106019	99190

* $p<0.05$, ** $p<0.01$, *** $p<0.001$; Numbers in parentheses are robust standard errors; Dependent variable is the log of adjusted hourly wage (2003 prices); Other independent variables that exist in all regressions but are not reported are as follows: Highest level of education(Bachelor's degree(baseline), Master's degree, Doctorate degree, and Professional degree), whether the highest degree is earned in the US, field of education, gender, marital status, has child(ren), whether English is the mother tongue, physical disability, place of birth, employment region, employer size (in terms of number of employees), employment sector, and survey year fixed effect; The data on which the model is run for each group of immigrants (based on the first visa on which migrated to the US), includes all native observations plus only that group of immigrants; PR stands for Permanent Residence visa.

of arrival are all very insignificant. As explained earlier, Borjas in almost all of his papers uninterruptedly emphasized on the importance of cohort effects and also the potential cohort bias in cross-sectional studies. So, it might be quite shocking at the first look to see such a so called important factor means nothing. In order to make sure that arrival cohorts have no explanatory power in my regressions, I even tested the explanatory power and significance of them in the auxiliary regressions of the main model for both mean and variance, and they were very insignificant in those regressions as well. Such a finding should not be very unexpected and shocking though after being given enough thought to it.

Borjas' (1985, 1994, 1999) reasoning for pointing out the "cohort effects" and challenging Chiswick's (1978) results was only one thing which still makes a lot of sense. The data that Chiswick (1978) had used included people from every type. People with different schooling levels (from no education to PhD), different skills, different English proficiencies, and so on, who migrated to the United States during different eras. Each group of those immigrants, who entered at a certain time period, could have had different levels of human capital and different qualities compared to groups who entered during another time interval. Borjas (1985) also found more recent cohorts to be different in terms of capability and skills, so challenged Chiswick (1978) results on assimilation rate. He claimed that part of what Chiswick (1978) interpreted as assimilation rate should be credited to the differences in cohorts' qualities.

It should be noted that there is a big difference between the data that most of previous studies use in their cross-sectional analysis compared to the one that I use in this paper. While almost all of the previous studies use a sample from the whole population which includes every type of immigrants with different qualities who came to the US or migrated to another countries over different periods, the NSCG data is a sample from a very elite population who is not only highly educated, but also has education in science and engineering fields. Consequently, it is not at all unrealistic to assume that different cohorts of arrival in this data have very similar packages of human capital when they arrive to the United States. If that is the case, obviously, there will be no cohort effect and as a result, no cohort bias in the calculated assimilation rates for such an elite group.

Another interesting result which is reported in table 2.5 is the significantly big wage premium that immigrants with different residency statuses have over US-born in every single one of the seven above mentioned groups and sub-groups. As explained earlier, according to the literature, the estimated coefficient on the indicator (dummy) variable of any immigrant status (if significant) represents the log earnings difference between US born (native) workers and that group of immigrants at entry. Although the model is controlling for all human-capital-related, socio-economic, and demographic characteristics available in the NSCG data (which is indeed a reach data and provides researchers with so much information that could not be found all at the same time in other datasets), the problem of not being able to capture and control for ability and motivations of individuals is still an issue.

Existence of "selectivity" among the immigrant group(s) could be an important justification for such a considerable wage gap between natives and immigrants. After all, those individuals who could go through the very competitive immigration and adaption process from leaving everything behind in their homelands and migrating to the United States to totally getting assimilated in the US economy and finding a job in the US labor market, should be very motivated and capable people. These individuals are definitely significantly above average and the NSCG dataset only includes such qualified individuals as its immigrant individuals. As can be seen in table 2.5, those immigrants who migrated to the US on study visa or work visa have higher premiums for all residency statuses compared to those who migrated on a permanent residence (immigrant) visa or as a dependent of another individual.

For instance, those who migrate as students (on a student visa) need to go through a hard and competitive route to get to the point of having a job in the US. First, they need to get graduated and earn their degrees from US schools which are on average more demanding and more serious than other schools elsewhere. Then they need to apply for jobs, and many of employers prefer to avoid the costly and risky process of sponsorship for those who are not at least permanent residents of the United States. So, a big number of people compete for comparably few job openings that accept applicants with no green card. If they are good enough to win the competition and lucky enough to get an offer, the next step will be getting a work visa. Work visas, such as H-1B, have an annual cap and visas are issued through a lottery and the chance of winning for those who hold postgraduate degrees from US schools is about one half. So, with such a low chance of getting to a proper employment, those who get there should be much better than the rest. That is why as could be noticed, those immigrants who have migrated to the US on either work visa or study visa have a much higher premium over natives than those who came as permanent residents or dependents and did not need to be as good as the former two groups of immigrants to be able to survive and eventually get a proper job.

Some datasets provide information about the job and salary of immigrants right before migration. Some other datasets provide information about some exam grades or rankings that could be used as a weak proxy of ability. Controlling for such variables can potentially reduce the selectivity bias in the estimated coefficients, but unfortunately, the NSCG data provides no information about immigrants before migration or any other information that could be used as a means to deal with this important and potentially misleading issue.

Some other noticeable results reported in table 2.5 are the wage gap at entry between natives and immigrants among men and women and also the difference between the three immigration statuses within each group/sub-group. Results show that the positive premiums for immigrant women (for all of the three statuses) compared to native women is almost half of the premium of male immigrants over male natives. Various reasons could cause such a difference between the two sexes. One of them could be that female immigrants might not be as much better than their native counterparts as male immigrants are. Another reason could be that immigrant women, compared to immigrant men, might be victims of some discrimination. There might exist even more reasons for this difference, but one cannot be sure unless more information is available.

Looking within every group/sub-group, it could be noticed that those immigrants who are currently citizens or permanent residents are having some positive premium over those immigrants who are working in the United States on some type of visa and have not gotten their green cards yet. Though, there is not a significant difference between citizen immigrants and permanent residents in all groups. This shows that having permanent residency of the US, as opposed to be a temporary resident, could be considered as an advantage in the job market. Maybe those who incur some costs to their employers at the beginning (for doing the sponsorship), get paid less when they start working.

One of the main goals of this chapter is to compute the convergence/divergence rate between natives and different groups of immigrants to see what happens to the earnings gap at entry with more years immigrants reside in the United States. In order to do so, I use the estimated coefficients of age and age-squared for natives, and age, age-squared, YSM, and YSM-squared for immigrants which are reported in table 2.5, and by putting the estimate along with the age averages in equation 2.3, I compute the convergence/divergence rates (CR) at various years after migration (1, 5, 10, 15, and 30 years after migration) for all the seven groups and sub-groups. Table 2.6 reports the CR's for the seven groups after certain years of residence in the United States by immigrants.

As can be noticed, within all groups, the CR has a decreasing tone, and gets smaller with

Table 2.6: Convergence rate for different groups of immigrants based on cross-section Results on pooled data

	All	**Men**	**Women**	**PR**	**Work***	**Study***	**Dependent***
CR for immigrants 1 year after arrival	0.00246	0.00126	0.00617	0.01186	-0.00341	0.00474	0.00171
CR for immigrants 5 years after arrival	0.00193	0.00088	0.00520	0.01013	-0.00346	0.00395	0.00156
CR for immigrants 10 years after arrival	0.00127	0.00040	0.00400	0.00798	-0.00351	0.00295	0.00136
CR for immigrants 15 years after arrival	0.00061	-0.00008	0.00279	0.00582	-0.00357	0.00195	0.00117
CR for immigrants 30 years after arrival	-0.00137	-0.00152	-0.00083	-0.00064	-0.00374	-0.00104	0.00057

Note: CR stands for Convergence Rate; All values in this table are derived from estimates reported in table 5 using equation (2.3); * Results are derived from estimates that are not all statistically significant at 5% or below.

more years immigrants reside in the United States. At the beginning of the way, CR takes positive values, but the rate gets smaller the bigger YSM becomes. After 10-20 years (depending on the group) it goes negative which means either the increase in immigrants' earnings is not as fast as those of natives anymore, or the decrease in immigrants' wages is happening with a faster pace compared to the decrease in native wages. Anyways, since there exists a considerable earnings gap between natives and immigrants at entry with a positive premium for immigrants, when the CR is taking positive values, it means that the earnings gap is becoming bigger and the wages of two groups are diverging. When after many years of residence in the United States it finally takes negative values, it means that the gap has started to close down, but soon after people get to the end of their working lives and so the premium will never disappear.

Comparing men and women, as reported in table 2.6, it is interesting to see immigrant women's wages diverge from the wages of native women at a faster rate than immigrant men's wages getting further from those of native men. Recalling the difference in wage gap at entry between groups of men and women, it is interesting to notice that female immigrants at the time of migrating to the United States have a lower (about half) premium over their native counterparts compared with male immigrants' premium. However, their wage gap gets bigger with a considerably higher pace than immigrant men's does. Women's divergence rate after 1 year from arrival is about 0.6%, while for men this rate is only about 0.1%. A huge difference between the CR's of the two groups remains the case throughout all the years they live in the United States. For instance, while immigrant men's rate gets to a negative number after about 15 years of residence in the US, immigrant women's wages are still diverging from those of native women at the rate of about 0.3%. The reason could be that the adaption process happens faster among immigrant women compared to men. The reason could also be that although immigrant women are no different from immigrant men in assimilation, native women's wages get increased at a lower pace than native men's wages do. There might also be other reasons to justify this difference.

Those immigrants who came to the United States for the first time on a permanent residence (immigrant) visa have the highest divergence rates (at different number of years after arrival) compared to other groups and sub-groups. Like the case of women, this sub-group of immigrants also have the lowest premium over natives at entry compared to other types of immigrants (based on their first entry visa type). Nevertheless, they have the highest rates at which their wages get further from wages of natives with similar characteristics and qualifications. The gap between wages of such immigrants and natives is being increased with a pace about 1.2% one year after arrival. This rate comes down with more time they live at the US, but even after 15 years they still get further from natives with the rate of 0.6%. Although the calculated rates for other groups of immigrants are not significant at 5% or below (due to at least one insignificant estimated element involved in the calculations), comparing the other three groups of immigrants to those who migrated on an immigration visa shows a gigantic difference. The other three groups have higher premiums at entry, but their CR's are not even comparable to that of immigrants who migrated as permanent residents. Obviously, those who come to the US on immigrant visa have an easier path in front of them in the US. They are eligible to do whatever a native can do with no legal restrictions. Therefore, their assimilation process can happen considerably faster and easier.

2.4.3 Longitudinal Multivariate Analysis

In this part of the paper, I exploit the most important advantage of the NSCG data which is its longitudinal characteristic. I use what is now a standard immigrant economic assimilation model in the literature for my analysis:

$$\begin{aligned} \log w_{jt} = {} & \gamma_0 + \boldsymbol{I_j \gamma_1} + \boldsymbol{X_{jt} \gamma_2} + \gamma_3 A_{jt} * Nat_j + \gamma_4 A_{jt}^2 * Nat_j \\ & + \gamma_5 A_{jt} * Imm_j + \gamma_6 A_{jt}^2 * Imm_j + \gamma_7 YSM_{jt} + \gamma_8 YSM_{jt}^2 + f_t + \delta_{jt} \end{aligned} \tag{2.5}$$

where j is the index for individual, t is the index of time, Imm (Nat) is a dummy variable taking one (zero) for immigrants and zero (one) for natives, and f_t is the time fixed effect (as described earlier). The rest of variables are same as before, and the interpretation of estimated coefficients and how to derive the convergence/divergence rates (CR) are as explained over the section 2.3.1.

I have included the cohort of arrival dummies for immigrants in the $\boldsymbol{X}$ vector to control for the difference between arrival cohorts in skills and quality[16], but as discussed before, there is a perfect collinearity among years since migration, year of observation, and the year of arrival (cohort). As a result, the estimated coefficients of all the three cannot be separately identified. In order to solve this identification problem, as discussed by Borjas (1999), some

[16] As explained earlier, immigrants are divided into 7 separate groups based on their arrival year: 1-those who entered the US before 1949; 2-between 1950 and 1959; 3-between 1960 and 1969; 4-between 1970 and 1979; 5-between 1980 and 1989; 6-between 1990 and 1999; and 7-those who arrived after year 2000.

type of restriction must be imposed. I resolve this identification problem by keeping the period effect (f_t) the same for both immigrants and natives (Borjas, 1994).

As discussed earlier, because of the perfect collinearity between age, years since migration, and age at arrival, or among age at arrival, years since migration, and year of observation, bringing age-at-arrival to the model needs an additional restriction such as forcing the same age-earnings profile for immigrants and natives, which is very restrictive and unrealistic. An alternative approach, as discussed, would be using a categorical variable for different age groups at arrival, which although common between researchers, is not a real solution to the problem[17].

To run the model on the panel data, I first assume that individual specific effects are uncorrelated with all of the independent variables. So, I use the random effects model to run the regression on the longitudinal data using generalized least square (GLS) estimation. Like the cross-sectional analysis, I run the random effects model four more times to compare the natives with different groups of immigrants based on their first entry visa on which they entered the United States.

As mentioned in the cross-section section, a known problem that has always been an issue when using wage equation, is the inability of the model to capture and control for ability and incentive of different individuals. Unfortunately, almost no dataset provides researchers with variables that can control for those attributes, while those unobserved characteristics of different individuals can affect their wages through productivity. Therefore, even after controlling for the maximum number of characteristics available in the data, any significant difference found between wages of different groups of people can still be (at least a part) because of differences in abilities and incentives.

If one assumes that individual and time fixed effects are dependent on one another, the ideal estimation approach to be used could be the fixed effects model. The fixed effects model captures all of the unobservable time-invariant characteristics such as ability and motivation. Such approach, however, removes any time-invariant variable (such as native/immigrant dummy, or gender dummy) from the model, and estimates of such variables' coefficients are usually of much interest to researchers. Nakamura and Nakamura (1985) introduce a model

[17] I also ran my longitudinal model with such a categorical variable (taking three values for those with less than 18 years of age at arrival, those between 18 and 30, and those above 30). Results showed that the estimated coefficients for age-at-arrival were not significant neither in the main model, nor in any of auxiliary regressions. Its inclusion did not change the other estimated coefficients either.

called “inertia model” for the above mentioned unobservables problem which is an alternative variation of the fixed effects model. They model individual fixed effects as a linear function of previous period’s log of real (annual) earnings and log of number of employed weeks in the year plus an error term. They argue that these lagged variables capture all individual-specific time-invariant characteristics that have the same impact on earnings of individuals in every year of their working lives. So, they claim that by adding these lagged terms, the effect of unobserved characteristics such as ability and incentive would be captured to a good extent. Since by bringing lagged income and lagged number of weeks worked to the right-hand side of the model coefficient estimates are capturing the effect of changes from the last survey cycle to the next, the inertia model has only been used (for instance in Chiswick et al. (2005a) and Gagliardi and Lemos (2015)) to calculate the wage gaps between natives and different groups of immigrants, and not for finding the assimilation/convergence rate.

Since by adding the lagged variables to the model one time point of the panel data will be gone, it is more preferred to use the inertia model when there are many points of time available in the longitudinal data. In this study, although I just have three cycles to use, I add the lagged variable of the log of hourly wage to the right-hand side of equation 2.5 and report the results with and without the lagged variable[18]. It should be highlighted that since I have already calculated the hourly wages of individuals in the data using their number of weeks worked per year and number of hours worked per week, I only add the lag of log hourly wage to the right-hand side of the model and report the results[19].

2.4.4 Longitudinal results

Table 2.7 demonstrates the results gained after running the random effects model (equation 2.5) on the longitudinal data of those individuals who are present in all the three cycles of NSCG (2003, 2006, and 2008)[20]. The left side of the table (columns (1) through (7)) shows the results when the model is run without the lagged income, and the right side of the table

[18] See table 2.7.

[19] I also ran the inertia model using the log of reported annual salary as the dependent variable and added lagged log of weeks worked per year and lagged log of hours worked per week along with lagged log of annual salary to the right-hand side of the model. The results were similar to those reported in table 2.7.

[20] Like the case for cross-section, in longitudinal analysis also the estimated coefficients for the dummies made for cohorts of arrival are very insignificant in the main model and in auxiliary regressions of the main model for both mean and variance. I even ran the RE model clustered by cohort of arrival dummies, but the results did not change significantly.

(columns (8) through (14)) shows the results with the lagged income added to the right-hand side of the model. Columns (1) and (8) report the results of running the model on the whole sample. Columns (2)&(9) and (3)&(10) show results for men and women on both models, respectively, and columns (4) through (7) and (11) through (14) demonstrate the longitudinal results after running the two models on the four following sub-sets of data: natives and immigrants who migrated on immigrant visa, natives and immigrants who migrated on work visa, natives and immigrants who migrated on study visa, and natives and immigrants who migrated on dependent visa. As discussed earlier, the model with lagged income is only used to find wage gaps between natives and immigrants in different groups, not for finding convergence/divergence rates.

Similar to the results gotten from cross-section model, longitudinal analysis also shows that within all seven groups, those immigrants whose current residency status is either naturalized citizen or permanent resident are having higher premiums over natives than those who are temporary residents of the United States. Though, an interesting difference between estimated premiums of longitudinal and cross-sectional approaches is that estimations of the longitudinal approach show a smaller wage gap between immigrants with different residency statuses and US-born in all groups. This means that cross-section approach over-estimates the earnings gap between natives and immigrants at entry. One way of explaining this difference could be through the concept of "survivor bias": Those immigrants who were attrited from the sample (for reasons other than those random and non-patterned ones that could cause attrition within both natives and immigrants similarly, such as death, opportunity cost of filling surveys, etc.) could have been more successful ones in the US labor market. They moved out of the US to either go back to their home countries to serve or for a better position, or go to another country as they received better offers outside the United States. In case such a scenario is true, those who stay are not as capable and qualified as those who left and this could cause the difference between estimations of the two approaches. Unfortunately, since the source(s) of attrition in the NSCG data is not known to the researcher, being sure about the reason of the aforementioned difference is not possible.

Although the estimated premiums with the longitudinal approach are a bit smaller than those found using cross-sectional approach, they are still showing a very considerable gap between the two groups. Since I control for all available characteristics of individuals in the model, like the cross-section case, only a strong selectivity bias among immigrants could justify such a difference. As discussed earlier, those individuals who successfully go through

Table 2.7: Longitudinal Results (Random Effects Model)

	Model without lagged income							Model with lagged income						
	All	Men	Women	PR	Work	Study	Dependent	All	Men	Women	PR	Work	Study	Dependent
Immigration Status														
Naturalized Citizen	0.617***	0.758***	0.372*	0.445**	0.991***	0.875***	0.236	0.475***	0.510***	0.354	0.292	0.243	0.770***	0.469
	(0.103)	(0.123)	(0.185)	(0.151)	(0.242)	(0.198)	(0.286)	(0.105)	(0.125)	(0.193)	(0.158)	(0.246)	(0.191)	(0.304)
Permanent Resident	0.630***	0.785***	0.351	0.434**	0.986***	0.881***	0.183	0.478***	0.514***	0.348	0.268	0.243	0.750***	0.445
	(0.102)	(0.122)	(0.184)	(0.150)	(0.239)	(0.196)	(0.283)	(0.104)	(0.124)	(0.192)	(0.158)	(0.244)	(0.189)	(0.300)
Temporary Resident	0.600***	0.751***	0.338	-	0.961***	0.827***	0.171	0.470***	0.499***	0.380*	-	0.195	0.732***	0.441
	(0.101)	(0.120)	(0.182)		(0.236)	(0.194)	(0.282)	(0.104)	(0.124)	(0.193)		(0.243)	(0.187)	(0.313)
Age (Natives)	0.053***	0.058***	0.048***	0.053***	0.053***	0.053***	0.053***	0.017***	0.017***	0.020***	0.016***	0.017***	0.017***	0.016***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Age (Immigrants)	0.030***	0.029***	0.035***	0.031***	0.022*	0.015	0.042***	-0.001	-0.002	0.005	0.002	0.010	-0.014	-0.004
	(0.004)	(0.005)	(0.007)	(0.006)	(0.011)	(0.009)	(0.013)	(0.004)	(0.005)	(0.008)	(0.006)	(0.011)	(0.008)	(0.013)
Age Squared (Natives)	-0.001***	-0.001***	-0.000***	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age Squared (Immigrants)	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000	-0.000**	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
YSM	0.009***	0.011***	0.007*	0.016***	0.004	0.018***	0.007	0.002	0.001	0.006*	0.007**	-0.001	0.005	0.005
	(0.002)	(0.002)	(0.003)	(0.002)	(0.005)	(0.004)	(0.004)	(0.001)	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)
YSM Squared	-0.000**	-0.000***	-0.000	-0.000***	-0.000	-0.000**	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-Squared within	0.149	0.165	0.128	0.145	0.149	0.149	0.149	0.020	0.013	0.030	0.024	0.021	0.022	0.022
R-Squared between	0.270	0.248	0.249	0.270	0.276	0.278	0.273	0.764	0.769	0.741	0.769	0.769	0.767	0.771
R-Squared overall	0.240	0.228	0.217	0.240	0.246	0.247	0.244	0.536	0.540	0.497	0.539	0.542	0.537	0.542
Number of Individuals	30913	19968	10945	26795	25529	26420	25060	30913	19968	10945	26795	25529	26420	25060

* $p<0.05$, ** $p<0.01$, *** $p<0.001$; numbers in parentheses are robust standard errors; coefficients are estimated using GLS with Random Effects Model assumptions; both of the models on the right and left of the table are allowing natives and immigrants to have different coefficients of age and age-squared; other independent variables that exist in all regressions but are not reported are as follows: Highest level of education (Bachelor's degree(baseline), Master's degree, Doctorate degree, and Professional degree), whether the highest degree is earned in the US, field of education, gender, marital status, has child(ren), whether English is the mother tongue, physical disability, place of birth, employment region, employer size (in terms of number of employees), employment sector, and survey year fixed effect. Model with lagged income also includes the log of hourly wage of the last time point for each individual.

the very competitive immigration and adaptation process (from leaving everything behind in their homelands and migrating to the United States, to totally getting assimilated in the US economy and finding a proper job in the US labor market), should be very motivated and capable people.

Except for the differences between magnitude of estimates for earnings gap at entry, every other discussion and comparison which was given in the section 2.4.2 for table 2.5 such as differences between men and women and also among the four different sub-groups of immigrants (based on their first entry visa type) holds true for longitudinal results (model without lagged income) as well.

As can be noticed at the right side of table 2.7, adding the last period's log of hourly income to the list of independent variables decreases the estimated wage gaps between natives and immigrants significantly. Although not all of the estimates are statistically significant, a considerable decline in the wage difference between the two groups is evident. This shows that even though the inertia model of Nakamuras captures the effect of unobservables (such as ability and motivation) on the wages, up to a point, it cannot completely capture the whole cause(s) of the wage difference between the two groups. Nevertheless, by looking at the R-Squared values which are reported for both models (without and with lagged income), it can be observed that there is a noticeable difference between the two models on all three reported R-Squared values (within, between, and overall). Going from the model without lagged log of income to the one with the lagged variable, the "R-Squared within" drops to somewhere around zero, while the "R-Squared between" and the "R-Squared overall" are drastically increased. This shows that the inertia model has been successful in capturing a big part of unobservables' effects on income.

As before, I use the estimated coefficients of age and age-squared for natives, and age, age-squared, YSM, and YSM-squared for immigrants which are reported in table 2.7, and by putting the estimates along with the age averages in equation 2.3, I compute the convergence/divergence rates (CR) for all the seven groups and sub-groups. Table 2.8 reports the computed CR's for the seven groups after certain years of immigrants' residence in the united states (after 1, 5, 10, 15, and 30 years).

Results shown in table 2.8 are in some ways very different from those of table 2.6. Still, within all groups, the CR show a decreasing manner, and gets smaller with more years immigrants reside in the United States. A positive CR at the beginning becomes negative

Table 2.8: Convergence rate for different groups of immigrants based on longitudinal results (model without lagged income)

	All	**Men**	**Women**	**PR**	**Work***	**Study**	**Dependent***
CR for immigrants 1 year after arrival	0.00355	0.00416	0.00366	0.00768	0.00035	0.01293	0.00488
CR for immigrants 5 years after arrival	0.00285	0.00332	0.00307	0.00642	0.00013	0.01085	0.00450
CR for immigrants 10 years after arrival	0.00198	0.00227	0.00234	0.00485	-0.00015	0.00824	0.00402
CR for immigrants 15 years after arrival	0.00110	0.00122	0.00161	0.00328	-0.00042	0.00564	0.00354
CR for immigrants 30 years after arrival	-0.00152	-0.00193	-0.00060	-0.00143	-0.00125	-0.00217	0.00210

Note: CR stands for Convergence Rate; All values in this table are derived from estimates reported in table 7 using equation (2.3); * Results are derived from estimates that are not all statistically significant at 5% or below.

after 10-20 years (depending on the group) of immigrants' residence in the US. As there exists an earnings gap between natives and immigrants at entry (with a positive premium for immigrants), while the CR is taking positive values, the earnings gap become wider and the wages of two groups get further. Since couple of years after it finally starts taking negative values people get retired, longitudinal approach also shows that the premium will never disappear.

Comparing men and women, as reported in table 2.8, unlike the cross-section results, the longitudinal approach shows that immigrant men start with a slightly higher rate, but their CR gets declined with a faster pace with more years they stay in the United States. Recall that, upon arrival, immigrant men have higher premiums over their native counterparts compared to immigrant women. With this result, such a difference between men and women will be kept the same and even increased. After about 10 years of residence in the US (by both genders of immigrants), finally the premium difference between the two groups start to decline very gradually. This could be interpreted as a sign of a persistent discrimination against women in the US job market. However, there could be other interpretations for it.

In contrast with cross-section results, longitudinal results show that those immigrants, who came to the United States for the first time on a student visa, have the highest divergence rates (at different number of years after arrival) compared to other sub-groups of immigrants. Recall that immigrants who have migrated to the United States on a study visa have the second highest estimated premium at entry compared to other sub-groups of immigrants. Table 2.8 shows that the gap between Wages of such immigrants and natives with similar characteristics and qualifications is being increased with a pace about 1.3% one year after arrival. Although this rate comes down with more time they live at the US, still after 15 years they are getting further from natives with the rate of 0.6%. The calculated rates for some of the other groups of immigrants (based on their first entry visa) are not significant at 5% or below due to at least one insignificant estimated element involved in the calculations.

Nevertheless, comparing the other three groups of immigrants to those who migrated on a student visa shows a gigantic difference. This shows that individuals who migrate to the US as students assimilate in the US economy considerably faster. This could be caused by their smartness and also higher cognitive abilities compared to other groups. Immigrants who migrated on an immigrant visa have the second highest divergence rates in table 2.8. The plausible reason for their high rates was discussed earlier. This is an interesting finding that results driven from the two approaches are showing different paths that different groups or sub-groups of immigrants traverse from the point they enter the United States of America.

2.5 Conclusion

A big part of immigration economics literature is devoted to studying the assimilation of immigrants in the labor market of the host country. Immigrants move to the host country with different human capital packages that might or might not match the type of human capital appreciated in the host country's labor market. The more they live in the host country, the more they invest on adapting themselves to the host country's economy, language, culture, and customs. Most of the previous studies use one or multiple cross-section(s) (depending on availability of data), and, unlike this study, find an earnings gap against immigrants at entry time as well as a rate at which immigrants catch up with natives and the gap is filled. It is argued in most of the cases that the earnings convergence is caused by the fact that immigrants are more able, more motivated or more hardworking than natives. However, most of these characteristics cannot be observed and controlled for.

Since Chiswick's 1978 paper, there have been debates on how to interpret findings gained from cross-sectional analysis. As one of the first critiques, Borjas (1985, 1995a) raised the "cohort bias" issue and emphasized that the cross-section comparison may not properly measure the true rate of assimilation, because there may be substantial differences in earnings potentials across year-of-arrival cohorts of immigrants. He showed that accounting for these cohort effects in wage levels, might change the rate of economic assimilation derived from cross-sectional approach crucially (Borjas, 2013). If cross-sectional estimates on assimilation mostly capture changes in unmeasured dimensions of immigrants' skills in different arrival cohorts, obviously, such results do not have much to say about the true assimilation process of immigrants. "Survivor bias" might also occur when multiple cross-sections from different years are used to find the assimilation or convergence rate. If return migration of immigrants that are less able and consequently less successful in the labor market occurs, cross-sectional results will over-estimate the real assimilation rate. However, if more able immigrants return to their home countries or migrate to another countries, the cross-sectional estimates will be under-estimating the real rate. As discussed by different scholars such as Chiswick (1980) and Borjas (1985), the ideal way of reaching accurate estimates and avoiding the above mentioned biases is to use a balanced longitudinal (panel) data which follows the same people over time, which is a rare dataset to find.

In this paper, using three cycles (2003, 2006, and 2008) of the reach and under-explored longitudinal national survey of college graduates (NSCG), I estimated the earnings gap at entry between highly educated natives and highly educated immigrants with different current residency statuses (naturalized citizens, permanent residents (Green Card holders), and visa holders) for seven groups and sub-groups: the whole sample, men only, women only, natives vs. immigrants who migrated as permanent residents, natives vs. immigrants who migrated on work visa, natives vs. immigrants who migrated on study visa, and finally natives vs. immigrants who migrated on dependent visa. I also computed the convergence/divergence rates for all the seven groups. To do so, I employed both cross-sectional and longitudinal approaches, and highlighted the differences between the results gained from the two approaches.

In contrast to almost all of the previous studies' results, both approaches show that immigrants with a bachelor's degree or higher have a huge premium over their native counterparts at entry. The more interesting result is that not only the wage gap at entry is positive for immigrants, but also this gap between natives and different groups of immigrants even get wider for the first 10-20 years of immigrants' residence in the United States. However, the initial gap and the rate at which earnings of natives and immigrants diverge are different between men and women and also between separate groups of immigrants based on their first entry US visa. Another considerable finding of this paper is non-existence of cohort bias, regardless of the approach taken. This shows that there is no significant difference between skills and quality of highly educated immigrants who entered the United States over different time spans (decade in my study).

Findings of this study show that the results gained from the two different approaches point to distinct routes that different groups and sub-groups of immigrants wend from the time they enter the US. Cross-section results show a higher premium for all groups of immigrants at entry compared to what longitudinal results show. Convergence/divergence rates calculated using these two approaches are also different. For instance, while cross-section shows immigrant women's wages get diverged from those of native women with a higher rate compared to immigrant men's rate of divergence, longitudinal results show otherwise. Or, while cross-section shows those immigrants who migrated to the US as permanent residents have the highest divergence rates compared to the rest of immigrants, longitudinal approach finds those who moved to the US on a student visa to have the highest rate among all.

As mentioned earlier, using balanced longitudinal data is supposed to help researchers with avoiding cohort bias and survival bias. My results from both approaches proved the former bias not to be the case among such educated sample. However, the survival bias can still be an issue when the cross-section approach is employed. Using a highly balanced longitudinal data can address that issue, but at the same time, the problem with longitudinal datasets, such as NSCG, especially when used for immigration-related studies, is the attrition. When the researcher is aware of the source of attrition, any bias caused by it could be controlled, but when the source is not known to the researcher, the results might not be very reliable. In case there exists a type of attrition with an unknown source that only affects some groups/types of individuals in the longitudinal data (like out-migration which is immigrant-specific), estimated coefficients and results derived from them could have upward or downward bias and provide the researcher with misleading results.

Other than the attrition problem, another issue that affects this study and other similar studies is existence of selectivity within some groups of individuals in the data. As explained earlier, immigrant observations of the data are representing a very special group of people. For the above explained reasons, an average immigrant with any country of origin is significantly more able and more motivated, than an average native of that country. Nevertheless, in studies like this one, these highly capable and motivated people are compared with a sample of natives of the host country which represents the whole population. Obviously, there is selectivity within immigrants, and, on average, they have higher abilities and motivations than natives of the host country. So, unless one can somehow take care of such a selectivity problem, the estimated results might not be very reliable.

Chapter 3

Respiratory Health Effects of $PM_{2.5}$ in Presence of Pollen: The Case of Reno/Sparks Metropolitan Area in Northern Nevada

3.1 Introduction

It is known to almost everyone that air pollution is an important cause of many respiratory illnesses and cardiovascular health problems. Air pollution could be caused by the existence (above certain levels) of chemicals such as Carbon Monoxide (CO), Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), and/or Ozone (O3) in the air. Presence of Atmospheric Particulate Matter (PM) of small diameters (especially $PM_{2.5}$) can be another source of severe air pollution (Dockery and Pope, 1994). For example, wildfire smoke is one of the main sources of $PM_{2.5}$, especially in areas with high frequency of wildfires. $PM_{2.5}$ has the potential to enter deeply into the lung when inhaled, and consequently cause short-term and long-term respiratory health problems in people of different ages. Kochi et al. (2012) estimate the cost associated with excess mortality as a result of being exposed to wildfire smoke during the 2003 southern California wildfires. They identify 133 excess cardio-respiratory related deaths caused by wildfire smoke exposure, controlling for factors such as seasonality and fluctuation of daily mortality levels. Based on their finding, they report the mean estimated total mortality related cost associated with the 2003 southern California wildfire to be around one billion U.S. dollars.

On the other hand, various types of pollen can also be the reason for respiratory health issues, especially among individuals with history of asthma (Erbas et al., 2007). There are papers in the literature that have studied the relationship between the level of PM's (mostly PM_{10}) and respiratory health or between the count of different kinds of pollen and respiratory problems, based on the number of hospital admissions (Jalaludin et al., 2004; Fan et al., 2016; Gleason et al., 2014). However, there is not much research done on the health impact of $PM_{2.5}$ in presence of pollen. The few studies which have addressed a similar research question, have mostly focused on asthma patients, especially children, and/or have studied the effect of PM_{10} rather than $PM_{2.5}$ (Jalaludin et al., 2000; Tobias et al., 2004; Jalaludin et al., 2004; Gleason et al., 2014).

This research takes place in the metropolitan area of Reno/Sparks, Nevada, and exploits a daily time series data from 2009 to 2015 for its analyses. Our objective in this paper is to find the true causal effect of $PM_{2.5}$ daily average (and also averages of the two previous days) on the number of in-patient, out-patient, and total respiratory admissions within all ages and also within three age groups: below 5 years old, between 5 and 65 years old, and above 65 years old, while controlling for pollen counts. Our main research question is to examine how important it is to control for pollen counts to obtain unbiased marginal effects of air pollutants (especially $PM_{2.5}$) on respiratory hospital admission counts/rates.

The study closest to our research is Moeltner et al. (2013). They study the effects of wildfire smoke on the number of in-patient admissions in Reno/Sparks, Nevada. However, they do not control for the effects of pollen, but rather use temporal indicators to control for pollen and other contemporaneous effects. This research is using a wider time span compared to their paper, and while controlling for pollen effects, identifies the effect of $PM_{2.5}$ on the number of both in-patient and out-patient respiratory admissions within different age groups.

In order to carry out our study in a way that it gives reliable coefficients for $PM_{2.5}$ and pollen, we control for a comprehensive group of air pollution and meteorological factors such as Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Ozone (O3), average daily temperature, min./max. daily temperature, average daily dew point, average wind speed, maximum wind speed, relative humidity, prescription, and average daily air pressure. We also control for the effects of months of the year, days of the week, and also national holidays. These variables are the common meteorological, air pollution, and temporal factors that are used as controlling variables in the literature, depending on their availability (Jalaludin et al., 2000; Tobias et al., 2004; Jalaludin et al., 2004; Gleason et al., 2014). Similar temporal variables are also used in the Moeltner et al. (2013)'s paper hoping that they will capture pollen effects without controlling for pollen count itself. However, it is not for certain whether or not these temporal effect are sufficient to capture the whole impact of pollen.

Our results depict a weak but significant positive relationship between the level of $PM_{2.5}$ and the number of total admissions and out-patient admissions among all patients and within all of the age groups except children under 5 years old[1]. However, there is no significant relationship between $PM_{2.5}$ and the number of in-patient admissions. Also, estimated coefficients for the lags of $PM_{2.5}$ are not significant within most of the age groups. Surprisingly, based on our results there is no evidence that shows any relationship between the pollen count and the number of total admissions or in-/out-patient admissions for any age groups. In order to check for the robustness of this result, we ran our model using three different specifications to control for pollen, but none of them led to a different result[2].

[1]One reason for not finding any significant relationship between $PM_{2.5}$ and the number of total admissions and out-patient admissions within children under 5 can be the smaller sample size for this age-group compared to the other groups. Another reason can be that such a very vulnerable group might react to $PM_{2.5}$ exposure by showing non-respiratory related symptoms

[2]Please see Appendices A, B, and C.

This chapter is organized as follows: Section 3.2 reviews the most important previous research done in the literature. In section 3.3 the data used for the study will be explained along with some descriptive statistics. Our methodological approach will be explained throughout section 3.4 and results will be discussed. Section 3.5 concludes.

3.2 Literature Review

Due to lack of daily data (especially in the United States), not many papers have studied the joint adverse impact of air pollutants and allergens on respiratory health. Air pollution caused by wildfires, industries, vehicles, etc. along with increases in the level of allergens such as pollen lead to many respiratory health issues among people of different age groups, especially individuals with asthma (Jalaludin et al., 2004; Fan et al., 2016; Gleason et al., 2014). Obviously, such health issues could be costly in different ways to the society. For instance, Kochi et al. (2012) estimate the cost associated with excess mortality as a result of being exposed to wildfire smoke during the 2003 southern California wildfires. They find that the mean estimated total mortality related cost associated with the 2003 southern California wildfire to be around one billion U.S. dollars. Most of the research conducted around this topic is studying other countries, rather than using US data. Most of the studies conducted in this capacity either do not discuss the implications of not controlling for pollen count when estimating pollution health effects; or do not disaggregate health effects by age groups and/or by in-patient/out-patient status; do not study the comparative effects of pollutants and pollen; or only focus on asthmatic health problems. Moreover, the majority of papers which study the health effects of pollutants and/or allergens do not study the adverse economic effects of the negative health impact. In this section, we review some of the papers which study the negative health effects of pollutants, pollen, or both. In order to cause a better understanding regarding the contribution of papers to the literature, they will be presented in chronological order.

Anderson et al. (1998) study the relationship between daily hospital admissions for asthma (all ages together and the age groups 0-14, 15-64 and 65+ years) and air pollution in London in 1987-1992 and the possible confounding and modifying effects of airborne pollen. They use a Poisson regression to estimate the relative risk of daily asthma admissions associated with changes in ozone, sulfur dioxide, nitrogen dioxide and particles (black smoke), controlling for time trends, seasonal factors, calendar effects, influenza epidemics, temperature, humidity, and auto-relationship of unobservables. In order to find independent effects of individual pollutants and also interactions with aero-allergens (three types of pollen), they use models with and without pollen count. They find that Ozone, sulfur dioxide, nitrogen dioxide, and particles all have significant relationship with daily hospital admissions for asthma. However, based on their results, associations with pollutants are not confounded by airborne pollens and there is no evidence that the effects of air pollutants and airborne pollen interact in causing hospital admissions for asthma. The estimated coefficients for pollen for different groups are not also showing any explanatory power.

Jalaludin et al. (2000) use daily air pollution, meteorological, pollen and alternaria (a type of fungus which is also a common allergens in humans) data, and study children with a history of wheeze to determine the associations between the Sydney's January 1994 bushfire period and PM_{10} and evening peak expiratory flow rates (PEFR is a test that measures how fast a person can exhale). They do not find an association between the bushfire period or PM_{10} and evening PEFR, although in a subgroup of children without bronchial hyper-reactivity, they find a significant negative relationship between PM_{10} and evening PEFR. No interaction term is used in their model, and the estimated coefficient for pollen is not reported.

Tobias et al. (2003) perform a time series analysis, adjusting for meteorological factors and air pollution variables, to study the short-term effects of different types of pollen on asthma hospital emergencies in the metropolitan area of Madrid (Spain) for the period of 1995-1998. They find a positive relationship between pollen levels and asthma related emergencies, independent of the effect of air pollutants. They use a Poisson model for their time series analysis, but they do not control for $PM_{2.5}$ as an important air pollution factor, and do not control for interactions between pollen and air pollution.

Tobias et al. (2004) investigate the potential non-linear short-term effects of different types of pollen with allergenic capacity on the daily number of asthma-related hospital emergency room admissions in Madrid (Spain) for the period 1995-1998. They use Poisson regression with generalized additive models, controlling for trend and seasonality, meteorological variables (mean temperature and relative humidity), acute respiratory infections, and air pollutants (PM_{10}, SO2, NO2, and O3). They categorize pollen into five groups defined on the basis of their respective distributions. They find that pollen with allergenic capacity in Madrid are positively related to asthma hospital emergencies. It should be noted that they have also controlled for lagged impacts of pollen.

Jalaludin et al. (2004) enroll a cohort of primary school children (125 students) with a history of wheeze in an 11-month longitudinal study to examine the relationship between ambient air pollution and respiratory morbidity. They use daily air pollution (ozone, PM_{10}, and nitrogen dioxide), meteorological, and pollen data. They exploit logistic regression models to determine relationship between air pollution and respiratory symptoms, asthma medication use, and doctor visits for asthma. They find no significant relationship between ambient ozone concentrations and respiratory symptoms, asthma medication use, and doctor visits for asthma. However, their research shows a positive relationship between PM_{10} concentrations and doctor visits for asthma and between NO2 concentration and wet cough in single-pollutant and multi-pollutant models. Based on what is reported in their paper, the interaction terms between air pollutants and total pollen are not significant. Estimated coefficients for pollen are not reported.

Hanigan and Johnston (2007) investigate the relation between daily average ambient pollen with hospital admissions for total respiratory diseases and for asthma, chronic obstructive pulmonary disease (COPD), and respiratory infections separately in Darwin, Australia, during the period from April 2004 to November 2005. They control for PM_{10}, average temperature, relative humidity, and precipitation with no interaction terms. Their findings show a relationship between pollen count and total respiratory hospital admissions, while there is no evidence for a relationship between pollen count and asthma admissions. They believe their results suggest that ambient airborne pollen might have a wider public health impact than previously recognized.

Erbas et al. (2007) examine the relationship between increasing ambient levels of grass pollen and asthma emergency department admissions in children. They also attempt to determine whether these relations (if any) are seen only after a thunderstorm, or whether grass pollen levels have a consistent influence on childhood asthma ED visits during the season. They conduct a short time series study (using a semi-parametric Poisson regression model) for asthma admissions to the emergency department among children in Melbourne, Victoria. They control for meteorological (daily maximum temperature, relative humidity, and precipitation) and air quality parameters (NO2 and O3) during the selected period in 2003 without using any interaction terms in their model. They find a clear relationship between increased risk of childhood asthma ED attendance and levels of ambient grass pollen below 20 grains/m3, independent of any impact of thunderstorm-associated asthma.

Moeltner et al. (2013) use information on emissions generated by 24 large wildfires in the Reno/Sparks area of Northern Nevada over a 4-year period (2005-2008). By relating the daily acreage burned by these wildfires to daily data on air pollutants and local hospital admissions, and also by using information on medical costs, they compute the per-acre health cost of different wildfires (based on the type and amount of fuel and also distance from the impact area).They find that even fires as far as 200-300 miles can cause an increase in patient counts in local hospitals. Their results also show that the marginal impact per acre burned has a negative relationship with distance from the fire and a positive relationship with the fuel load. This is the closest paper to our study in the way that they study the exact same area. They use daily data to find the impact of pollution on health outcomes, but they do not control for the potential respiratory health impacts caused by pollen. Instead, they use temporal (monthly) indicators to control for the pollen effect. Of course, if pollen peaks move around across the year, this would be insufficient. In our paper we attempt to figure out whether temporal indicators are sufficient to capture the pollen effects or daily pollen count should be controlled for separately.

Gleason et al. (2014) study the temporary impact of ozone, $PM_{2.5}$, and pollen on acute onset of asthma in children in New Jersey. They use daily emergency department visits for children aged 3 to 17 years with a primary diagnosis of asthma during the warm season (April through September) of 2004-2007. They also control for lag exposures along with controlling for holidays, school-in-session indicator, temperature and relative humidity, gender, race, ethnicity and socio-economic status. They find that the ambient air pollutant ozone is correlated with increases in pediatric emergency department (ED) asthma visits during the warm weather season. Also, the different pollen types show different relationships with the emergency department admissions. Moreover, they find that high levels of tree pollen appear to be an important risk factor in asthma exacerbations.

Fan et al. (2016) study the impact of $PM_{2.5}$ on asthma emergency department visits, controlling for meteorological variables, demographic variables, pollen and holidays or weekends. The authors use a random-effects model to estimate the relationships. Their results show that with respect to short-term effects, asthma ED visits increase at higher $PM_{2.5}$ concentrations and children are more susceptible than adults to increased $PM_{2.5}$. They find that the ED visits increase more during the warm season per 10 g/m3 increase in $PM_{2.5}$ than during cold season. Results show that $PM_{2.5}$ has an adverse impact on asthma ED visits after short-term exposure and that children are a high-risk population when $PM_{2.5}$ concentrations are high, particularly in warm seasons. They do not report the estimated coefficients on their controlling variables and do not seem to have used any interaction terms between pollen and $PM_{2.5}$ in their model.

Other than the papers mentioned above, there are papers such as Butry et al. (2001), De Mendonça et al. (2006), Rittmaster et al. (2006), Vedal and Dutton (2006), Delfino et al. (2008), Jayachandran (2009), Kochi et al. (2012), and Shaposhnikov et al. (2014) that study the adverse impacts of wildfire smoke on the human health in different countries based on previous extensive wildfires in different areas. There are also some other papers such as Dennekamp and Abramson (2011), De Sario et al. (2013), Youssouf et al. (2014), and Liu et al. (2015) that perform a systematic review on studies linking wildfire smoke or other types of air pollutants or allergens and human health in different areas.

In this paper we study the separate impacts of $PM_{2.5}$ and pollen on the total, in-patient, and out-patient respiratory hospital admissions for all patients and also within different age groups using a long period of daily time series data. Papers already published in this literature are either focused on asthma patients or a specific age group, and are not separating in-patient from out-patient admissions in their studies.

3.3 Data and Descriptive Statistics

3.3.1 Hospital Admission Data

As explained earlier, the time frame of our analysis ranges from the beginning of 2009 to the end of 2015. Since the center collecting the pollen count data, which is the source of our pollen data, does not have readings from the beginning of November of each year to the end of February of the next year, we limited our study to March of each year through October. Hence, the descriptive statistics tables given below show the summaries and trends of hospital admission data, meteorological data, air pollution data, and pollen data for March to October of 2009 to 2015.

We obtained our data from different sources. Patient admissions data were provided by the Nevada Center for Health Statistics and Informatics at the University of Nevada, Reno. Under state law, this center collects and maintains billing records from Nevada hospitals and ambulatory surgical centers. Among other information, these medical systems are required to submit daily in-patient and out-patient data to this center on a regular basis. We consider all unscheduled in-patient and out-patients visits related to respiratory issues as captured by International Disease Codes (IDCs) 460.0-486.99, and 488.0-519.99, except for influenza (IDC 487.00-487.99) from Jan. 1, 2009 to Dec. 31, 2015. Each patient receives a primary ICD code upon admission. To be counted as an out-patient admission as opposed to in-patient, patient needs to be discharged within 24 hours. Otherwise, the admission will be considered as an in-patient (Moeltner et al., 2013).

The health impacts of air pollution and pollen might vary greatly over different age groups. In order to study any potential age-based differences, we divide patients into three age groups of “under 5”, “between 5 and 65”, and “over 65” and study each group separately. Table 3.1 displays respiratory admission counts for out-patient, in-patient, and total for the mentioned age groups. Overall, 67,986 patients were admitted for respiratory health problems during our research period, which is equal to an average of 39.64 admissions per day with a standard deviation of 12.72. The majority of this number comes from the out-patient admissions, which comprises 54,914 (on average 32.02 per day with standard deviation of 11.24) visits compared to 13,072 (on average 7.65 per day with standard deviation of 3.33) in-patient admissions. As evident in the table 3.1, among out-patient visits, age group of 5 to 65 has the biggest share, while patients over 65 years old have the larger share of in-patient admissions.

3.3.2 Meteorological Data

We downloaded the daily climate data from the National Centers for Environmental Information (NCEI)'s data repository website[3]. Summary statistics of the meteorological variables which are used in our models are given in table 3.2. According to the literature, all of the factors that can affect the amount and density of air pollutants or allergens, such as temperature, dew point, wind, humidity, air pressure, and precipitation are controlled for in our econometric model (Tobias et al., 2004; Jalaludin et al., 2004; Moeltner et al., 2013; Gleason et al., 2014; Fan et al., 2016). As described in table 3.2, the average daily temperature for our period of research (months March to October of years 2009 to 2015) is 63.19 degrees Fahrenheit with a standard deviation of 12.73 degrees. The average daily dew point for the same period is 30.53 degrees Fahrenheit with a standard deviation of 8.92 degrees. Wind has

[3]https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets

Table 3.1: Respiratory Patient Counts, Reno, March-October 2009-2015

Age Group	**Admission Counts**			
	Total	**Daily**		
		Mean	Std	Median
Out-Patient				
Under 5	13,847	8.07	4.44	7
5-65	37,427	21.82	7.91	21
Over 65	3,640	2.12	1.52	2
All	54,914	32.02	11.24	31
In-Patient				
Under 5	1,043	0.61	0.88	0
5-65	5,542	3.24	1.94	3
Over 65	6,487	3.80	2.16	4
All	13,072	7.65	3.33	7
All Admissions				
Under 5	14,849	8.69	4.74	8
5-65	42,854	25.08	8.46	24
Over 65	10,111	5.92	2.73	6
All	67,986	39.64	12.72	38

been blowing at the average daily speed of 5.75 knots (with a standard deviation of 2.97) with a humidity at a mean of 34.69% and standard deviation of 13.33%. As can be expected, not much precipitation is recorded throughout the period, and on average daily air pressure is 865.83 millibars (with a 3.81 standard deviation).

Table 3.2: Climate Data, Reno, March-October 2009-2015

Climate Variable	**Unit**	**Mean**	**Std**	**Min**	**Max**
Avg. Temperature	Fahrenheit	63.19	12.73	28.2	89.3
Min. Temperature	Fahrenheit	47.27	10.96	17.1	70
Max. Temperature	Fahrenheit	80.31	13.77	42.10	105.1
Avg. Daily Dew Point	Fahrenheit	30.53	8.92	1.2	56.8
Avg. Wind Speed	Knots	5.75	2.97	0.5	24.1
Max. Sustained Wind Speed	Knots	16.55	5.66	4.1	42
Relative Humidity	Percent	34.69	13.33	12.32	88.30
Precipitation	Inches	0.01	0.08	0	1.45
Avg. Daily Air Pressure	Millibars	865.83	3.81	848.5	878.2

3.3.3 Air Pollution Data

The daily/hourly data on different air pollutants including $PM_{2.5}$, Carbon Monoxide (CO), Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), and Ozone (O3) were downloaded from the AirData page of the United States Environmental Protection Agency (EPA)'s website[4]. Table 3.3 presents the summary statistics of the air pollution variables which are used in our analysis as control variables during the research period (March-October, 2009-2015). The main purpose of this research is to estimate the true relationship between air pollution factors (especially $PM_{2.5}$) and the number of hospital admissions. As reported in the table, for the period of research, we see a mean of 6.93 micrograms per cubic meter with a standard deviation of 6.83 for the level of $PM_{2.5}$. Also, the averages of carbon monoxide, Nitrogen dioxide, and ozone for the same period are 0.22 parts per million, 11.35 parts per billion, and 0.04 parts per million (with standard deviations of 0.07, 5.08, and 0.01), respectively.

Table 3.3: Air Quality/Allergen Data, Reno, March-October 2009-2015

Air Quality/Allergen Variable	**Unit**	**Mean**	**Std**	**Min**	**Max**
$PM_{2.5}$	Micrograms/Cubic Meter	6.93	6.83	1	100.96
Pollen	Count	135.41	192.22	1.80	1623.80
CO (Carbon Monoxide)	Parts/Million	0.22	0.07	0.04	0.88
NO2 (Nitrogen Dioxide)	Parts/Billion	11.35	5.08	2.07	36.13
O3 (Ozone)	Parts/Million	0.04	0.01	0.007	0.06

3.3.4 Pollen Data

Unlike many other countries which are collecting daily counts of different types of pollen, it is not easy in the United States to obtain daily data on pollen. Only certain cities in the US are collecting daily data on pollen and spores[5]. In order to obtain data on pollen in Reno, Nevada, we contacted Allergy and Asthma Associates of Reno, who are is listed on the website of the American Academy of Allergy Asthma and Immunology (AAAAI) as a certified local counting station and a member of the National Allergy Bureau[6]. Allergy and Asthma Associates voluntarily collects and reports counts of pollen in Reno. However,

[4]https://aqs.epa.gov/api

[5]List of the cities with collection centers can be found on the web page of Notional Allergy Bureau at http://www.aaaai.org/global/nab-pollen-counts?ipb=1

[6]http://pollen.aaaai.org/nab/index.cfm?p=DisplayStationInfo&stationid=86

the data is neither collected daily, nor with a constant pattern/frequency. Sometimes they do not collect it for up to 10 days and sometimes they collect data in two days in a row. On average, the data is collected once every seven to ten days. Also, there is no collection from November through February. Because of this shortcoming in the data, as mentioned earlier, we limited our study to March of each year through October. Table 3.3 shows the daily mean, standard deviation, minimum, and maximum values of pollen for the research period using the data received from the Allergy and Asthma Associates. The mean of daily pollen count for the period of our study is 135.41 with a standard deviation of 192.22. The minimum of all times during this period is 1.80 and the maximum (occurred in April 2009) is 1623.80.

3.3.5 Time-Series Graphs

The entire time series of daily admissions for respiratory illnesses (out-patient, in-patient, and total) is plotted in figure 3.1. Figure 3.2 demonstrates the graphs for total admissions, $PM_{2.5}$ and pollen count. In should be noted that in order to give a better understanding of the whole picture of the story, the graphs presented in figures 3.1 and 3.2 include the observations for all of the 12 months of each year from 2009 to 2015. Since lines in each graph are connecting each observation to the next (in chronological order), the long lines in the pollen graph (the last graph of figure 3.2) present the period of November to February, in which no pollen count data is collected.

As can be seen in figure 3.1, a clear seasonal pattern exists in in-patient, out-patient, and total admissions with peaks in late winter or early spring and troughs in late summer or early fall. The peaks are likely reflecting the poor air quality during the winter season due to inversion, combined with allergies that start in spring (Moeltner et al., 2013).

The second and third plots of figure 3.2 show the time-series of $PM_{2.5}$ and pollen, respectively. Several factors such as wildfires, house fires, traffic, etc. affect the level of $PM_{2.5}$. It seems like $PM_{2.5}$ level is following a systematic pattern. We see sharp peaks mostly around the first couple of months of each year. However, the unexpected jumps in the level of $PM_{2.5}$ outside of the routine peak months most likely occur because of some sudden and unexpected incidents like fires/wildfires around the recording station(s). Pollen level has a predictable pattern as well, which is not surprising. Climate of each region/area has an important role in determining when trees and grasses release their pollen each year. Also, anytime of the

Figure 3.1: Time-Series Plots for In-Patient Admissions, Out-Patient Admissions, and Total Admissions - Reno - 2009-2015

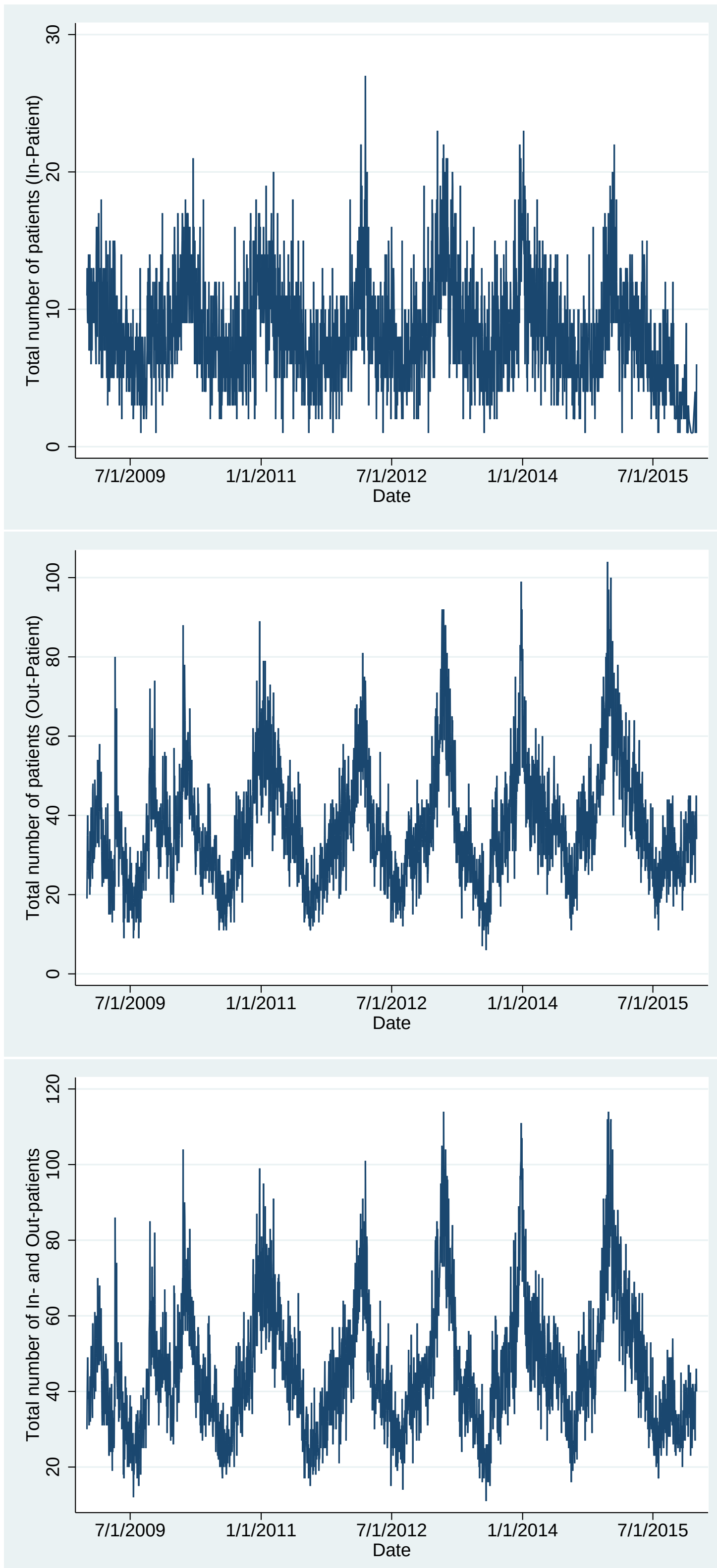
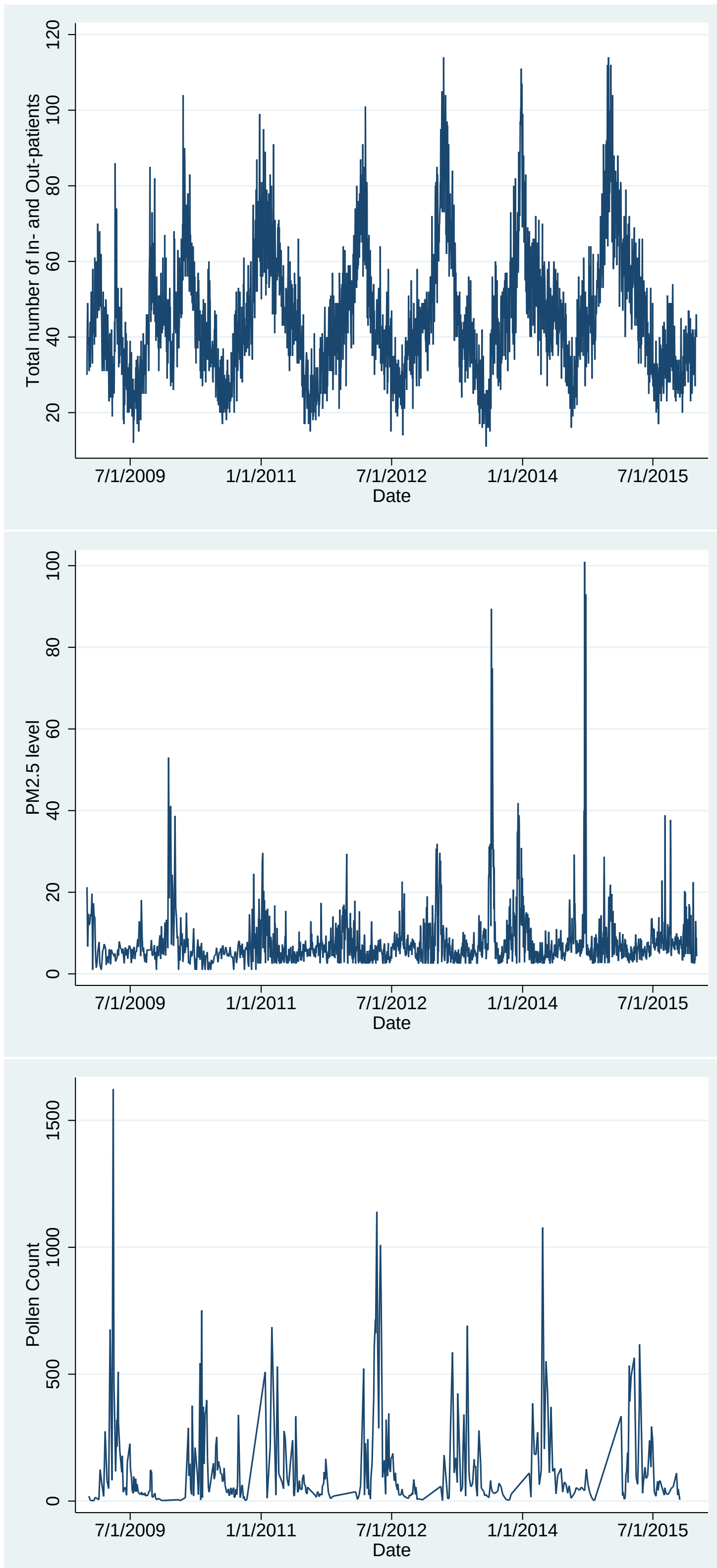


Figure 3.2: Time-Series Plots for Total Admissions, $PM_{2.5}$ Level, and Pollen Count - Reno - 2009-2015



year there are at last some plants which are releasing some pollen, but March to July is known as the period with maximum pollen release (Subiza et al., 1995).

By looking at all of the plots at the same time a mild relationships between the peak of admissions and that of $PM_{2.5}$ and also between the peak of admissions and that of pollen seem to exist, but although there seems to be some visual relationships, a more rigorous econometric approach is needed to identify causal effects.

3.4 Methodology and Results

3.4.1 Model Specification

This paper follows the existing contributions to the air pollution/allergens/health outcome literature[7] and models the effects of $PM_{2.5}$ and pollen on the daily number of respiratory hospital admissions (in-patient, out-patient, and total) for different age groups within a count data framework. Moeltner et al. (2013) considered several candidate specifications such as a standard Poisson model, a Poisson model allowing for over- or under-dispersion (also known as Poisson GLM model), a fully robust Poisson model with respect to any variance mis-specification (estimated via Quasi-Maximum Likelihood Estimator (QMLE)), and two Negative Binomial count models: the *NEGBIN I* and *NEGBIN II* specifications. They checked for over-dispersion and found that both NEGBIN models and the Poisson GLM point at very mild over-dispersion. Since it was difficult to further distinguish between these three closely related models, they chose the fully robust Poisson QMLE approach as their preferred specification for their patient model.

Using a similar analogy, we decided to use the fully robust Poisson QMLE approach for our analysis. As it is known, this approach produces consistent coefficient estimates under any variance mis-specification, as long as the conditional mean function is correctly specified (Moeltner et al., 2013). The model is given as:

(3.1) $f(y_t|\lambda_t) = \frac{exp(-\lambda_t).\lambda_t^{y_t}}{y_t!}$, where
(3.2) $E(y_t) = V(y_t) = \lambda_t$ and
(3.3) $\lambda_t = exp(\mathbf{z}'_p\boldsymbol{\beta} + \mathbf{z}'_m\boldsymbol{\gamma} + \mathbf{z}'_t\boldsymbol{\delta})$,

[7] Please see section 3.2.

where $f(y_t|\lambda_t)$ is the Poisson distribution function for the number of hospital admissions with the mean of λ_t. The parameterized mean function of our Poisson λ_t includes three sets of regressors: 1) $\mathbf{z}_p$, which is a vector of air quality/allergen indicators including the weekly average of pollen count, daily average level of $PM_{2.5}$ and its two lags for the two previous days, and the daily averages of Carbon Monoxide (CO), Nitrogen Dioxide (NO2), and Ozone (O3); 2) $\mathbf{z}_m$, which is a vector of meteorological indicators and includes average daily temperature, min./max. daily temperature, average daily dew point, average wind speed, max. wind speed, average daily relative humidity, daily prescription, and average air pressure of the day; and finally, 3) $\mathbf{z}_t$, a vector of temporal factors, including indicators for calendar months (March through October with March being the baseline), day of week (with baseline being Sunday), and indicator of national holidays.

Since the pollen data received from Allergy and Asthma Associates of Reno is not gathered on a daily basis (on average once every 7 to 10 days), we decided to use the weekly average of the existing pollen data as a proxy for daily pollen count. Using the weekly average to control for pollen also helps with capturing the lagged respiratory health effects of pollen. However, in order to check for the robustness of our results using weekly average of pollen count, we ran our model using 3 other specifications for pollen as well: 1) daily pollen count is used, and linear interpolation technique is hired to fill the missing data between two consecutive readings; 2) daily pollen count is used, and step function technique is employed to fill the missing data between two consecutive readings; and 3) only those days that have pollen count readings are used for the analysis. The results of running the model using these 3 approaches towards pollen are reported in Appendix A, Appendix B, and Appendix C, respectively.

Regarding $PM_{2.5}$, since it might be the case that respiratory symptoms do not become visible right away after being exposed to high levels of $PM_{2.5}$, and sometimes it might take couple of days before people realize they have serious respiratory problems and go to a hospital for it, we decided to control for two lags of $PM_{2.5}$ to make sure that in case a lagged effect exists, it will be controlled for.

As mentioned earlier, only those days which fall into the period of March through October are considered for our analysis, due to no collection of pollen count data outside this period. Also, we estimate separate admission count models for the entire sample, out-patients only, and in-patients only. Within each of these three, we run the model separately for the whole sample and also for the three age groups mentioned in section 3.3.1.

3.4.2 Estimation Results

As discussed earlier, our model is estimated via Quasi-Maximum Likelihood (QMLE) which generates fully robust standard errors for the coefficients in the parameterized mean function. Table 3.4 shows the estimation results of running this model for all respiratory admissions (both in-patient and out-patient), and the sub-groups of "under 5", "between 5 and 65", and "over 65". As can be seen in the table, $PM_{2.5}$ has a significant effect on the number of total respiratory admissions for all patients, those between 5 and 65, and also for those over 65 years old. For each 1 $\mu g/m^3$ increase in the $PM_{2.5}$, the model estimates increases in the expected number of respiratory admissions by 0.5%, 0.6%, and 0.6% among all patients, those between 5 and 65, and those over 65 years old, respectively. The first lag of $PM_{2.5}$ also has a slightly significant positive effect on the number of admissions of the same age groups (0.2%, 0.3%, and 0.5%, respectively), but the second lag does not seem to have any significant impact on the total admissions.

Pollen count, on the other hand, is not showing any effect on the number of admissions among any group or sub-group[8]. As explained in section 3.4.1, in order to test the robustness of our results, we ran our model using three other specifications for pollen. Tables A.1, B.1, and C.1 in appendices A, B, and C also report no pollen effect on the number of admissions among any group or sub-group, when the other approaches are taken. It should be emphasized though that we cannot be sure whether or not this means that pollen in general does not have any explanatory power here. It might be the case that our pollen data with many missing days cannot help us obtain reliable coefficients.

Carbon monoxide (CO) shows a significant and strong negative impact on the number of total respiratory admissions among all of the age groups. According to the United States Environmental Protection Agency (EPA)[9], being exposed to an increased level of CO will cause heart, lung, and brain issues that in most of the cases need in-patient hospitalization.

---

[8]The econometric results reported in our tables only go down up to three decimals, and the first three decimals of the pollen estimates were zeros. So, we decided to divide the pollen count by 100 first and then use it in the model.

[9]https://www.epa.gov/indoor-air-quality-iaq/carbon-monoxides-impact-indoor-air-quality

Table 3.4: Results for the total admissions (both in-patient and out-patient) - weekly average of pollen is used[1]

Variable	**All Patients**	**Under 5**	**Between 5 and 65**	**Over 65**
$PM_{2.5}$	0.005***	0.002	0.006***	0.006**
$PM_{2.5}$ (first lag)	0.002*	-0.001	0.003**	0.005**
$PM_{2.5}$ (second lag)	-0.001	0.001	-0.001	-0.005
Pollen (weekly average[1]/100)	-0.005	-0.008	-0.003	-0.007
CO	-0.405***	-0.461*	-0.314**	-0.792***
NO2	0.006**	0.009*	0.006*	0.001
O3	-0.989	-0.476	-1.387	-0.448
Avg. temperature	0.004	0.004	0.007*	-0.002
Dew point	-0.001	0.004	-0.002	-0.003
Max. temperature	-0.002	-0.000	-0.004*	0.000
Min. temperature	-0.005**	-0.014***	-0.003	0.002
Relative humidity	-0.001	-0.005	0.000	0.001
Avg. wind speed	0.004	0.017**	-0.000	-0.001
Max. wind speed	0.001	0.000	0.002	-0.000
Precipitation	0.052	-0.050	0.092	-0.061
Avg. air pressure	-0.004**	-0.004	-0.005*	-0.003
April	-0.246***	-0.316***	-0.233***	-0.184***
May	-0.209***	-0.316***	-0.174***	-0.188***
June	-0.386***	-0.523***	-0.368***	-0.259***
July	-0.647***	-0.924***	-0.617***	-0.422***
August	-0.600***	-0.792***	-0.541***	-0.560***
September	-0.320***	-0.473***	-0.234***	-0.465***
October	-0.242***	-0.307***	-0.196***	-0.320***
Monday	0.053**	-0.106***	0.088***	0.135***
Tuesday	0.001	-0.208***	0.062**	0.042
Wednesday	-0.063***	-0.204***	-0.019	-0.042
Thursday	-0.091***	-0.248***	-0.065**	0.029
Friday	-0.116***	-0.210***	-0.121***	0.038
Saturday	-0.088***	-0.154***	-0.073***	-0.056
National holiday	0.012	-0.044	0.041	-0.050
Constant	8.099***	6.755**	7.785***	5.038
Number of Observations	1511	1507	1507	1507

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

[1] Available readings of pollen count for the current day and the 6 previous days is used to calculate the weekly average of pollen. Then the result is divided by 100.

Our counter-intuitive negative estimates of CO coefficients might be due to one or more of the following reasons: First, according to the EPA, an increase in the CO level in most of the cases causes non-respiratory problems. So, when the CO level in the air is increased, due to the severity of non-respiratory health issues caused, less serious respiratory problems resulted from air pollution might get overshadowed. Secondly, since high levels of CO cause severe health issues, out-patient admissions, that would have happened at a lower CO level and due to other air pollution factors, might get replaced by in-patient admissions under high CO levels[10]. Since about 80% of the total number of admissions comes from out-patient admissions, the outcome can be expected to show a negative relationship between the total number of admissions and the CO level. Finally, in case the high CO level is reported over the public media, people would be advised to stay indoors. So, we might notice an averting behavior. Some individuals (potentially vulnerable ones) may choose to avoid getting exposed to a high level of CO by staying home or leaving town.

Results show a positive relationship between nitrogen dioxide (NO2) and the number of total admissions. For each 1 part per billion increase in the level of NO2, the model estimates increases in the expected number of respiratory admissions by 0.6%, 0.9%, and 0.6% among all patients, kids under 5, and those between 5 and 65 years old, respectively.

Most of the other air quality and meteorological factors are not showing significant effects on admission counts within our four groups/sub-groups. Month indicators and day-of-week indicators, on the contrary, show very significant effects on the number of total hospital admissions among all the age groups. Compared to March (baseline), all other months through October estimate fewer number of admissions. Likewise, any day of the week is expected to have less total admissions being compared with Sunday which is our baseline. This might be due to selectivity. People who work during the week might find Sunday to be the only day that they can give a visit to a hospital to take care of their respiratory issues. Finally, national holiday indicator shows no relationship with the number of admissions.

When we focus on out-patient admissions, the results are very similar to those of all admissions[11]. Table 3.5 captures the estimation results of running our model for out-patient admissions only, and again the sub-groups of "under 5", "between 5 and 65", and "over 65". As can be seen, $PM_{2.5}$ shows a significant effect on the number of out-patient respiratory

[10] As can be seen in table 3.6, a not very significant but positive relationship exists between the number of in-patient admissions and the CO level within most of the age groups.

[11] It is not much unexpected though, since about 80% of the total number of admissions comes from out-patient admissions.

admissions for all patients, those between 5 and 65, and also those over 65 years old. For each 1 $\mu g/m^3$ increase in the $PM_{2.5}$, the model estimates increases in the expected number of out-patient respiratory admissions by 0.7%, 0.7%, and 1.1% among all patients, those between 5 and 65, and those over 65 years old, respectively. The two lags of $PM_{2.5}$ prove to have no significant effects on the number of admissions within most of age groups. However, the first lag shows a significant relationship among the last two age groups (0.2% and 0.9%, respectively).

Analogous to the case of total admissions, pollen levels are not showing any effect on the number of admissions among any of the sub-groups. This is similar to the results gained by re-running our model using the other 3 proxies for pollen count. As reported in tables A.2, B.2, and C.2 (in appendices A, B, and C, respectively), the estimated pollen effect is insignificant in almost all of the groups/subgroups. In few subgroups the estimated coefficient of pollen is significant, but it is so small that it can be neglected.

While most of the air quality and meteorological factors show no relationship with the number of out-patient admissions, the model estimates a negative coefficient for CO within most the age groups which is also very significant. This negative impact is bigger among older people with ages over 65. As explained earlier, the reason can be that a bit of increase in the amount of CO might cause severe health issues that lead to in-patient admissions in respiratory and non-respiratory departments[12]. The model also estimates positive coefficients for NO2 within all patients and those between 5 and 65. Based on the results, for each 1 part per billion increase in the level of NO2, there will be increases in the expected number of out-patient respiratory admissions by 0.6% among all patients and those between 5 and 65 years old.

Average daily temperature seems to have a positive relationship with the number of out-patient admissions within all patients and patients between 5 and 65 years old. We expect to have on average about 1% increase in the number of admissions in return to 1 degree Fahrenheit increase in the average daily temperature. Minimum daily temperature also shows a significant relationship with the number of out-patient admissions within all patients and kids under 5. The model estimates a 0.7% decrease in the number of out-patient admissions within all patients in return to 1 degree Fahrenheit increase in the minimum daily temperature, while those patients under 5 years old will experience 1.6% decrease in the number of admissions for 1 degree Fahrenheit increase in the minimum daily temperature. The intuition

[12]Table 3.6 shows a not very significant but positive relationship between the number of in-patient admissions and the CO level within most of the age groups which can support this hypothesis.

Table 3.5: Results for out-patient admissions - weekly average of pollen is used[1]

Variable	All Patients	Under 5	Between 5 and 65	Over 65
$PM_{2.5}$	0.007***	0.003	0.007***	0.011**
$PM_{2.5}$ (first lag)	0.002	-0.002	0.002*	0.009**
$PM_{2.5}$ (second lag)	0.000	0.002	0.000	-0.007
Pollen (weekly average[1]/100)	-0.005	-0.009	-0.004	-0.009
CO	-0.534***	-0.562**	-0.449***	-1.292***
NO2	0.006**	0.008	0.006*	0.004
O3	-1.587	-0.203	-2.591*	3.980
Avg. temperature	0.009**	0.005	0.011**	0.005
Dew point	-0.002	0.004	-0.005	0.000
Max. temperature	-0.004*	-0.000	-0.005**	-0.007
Min. temperature	-0.007***	-0.016***	-0.004	-0.001
Relative humidity	-0.000	-0.005	0.002	-0.002
Avg. wind speed	0.005	0.017**	0.001	0.001
Max. wind speed	0.001	0.000	0.002	-0.010*
Precipitation	0.052	-0.037	0.078	0.030
Avg. air pressure	-0.003	-0.003	-0.003	-0.005
April	-0.252***	-0.319***	-0.233***	-0.163**
May	-0.198***	-0.288***	-0.164***	-0.167**
June	-0.404***	-0.492***	-0.377***	-0.329***
July	-0.688***	-0.891***	-0.645***	-0.423***
August	-0.599***	-0.758***	-0.538***	-0.615***
September	-0.296***	-0.429***	-0.228***	-0.490***
October	-0.217***	-0.244***	-0.201***	-0.262***
Monday	0.012	-0.138***	0.071**	0.004
Tuesday	-0.038	-0.231***	0.038	-0.076
Wednesday	-0.082***	-0.217***	-0.025	-0.124*
Thursday	-0.135***	-0.266***	-0.086***	-0.115
Friday	-0.162***	-0.221***	-0.147***	-0.065
Saturday	-0.106***	-0.159***	-0.088***	-0.079
National holiday	0.046	-0.022	0.066	0.051
Constant	6.968***	5.373*	6.479***	6.147
Number of Observations	1511	1511	1511	1511

* p<0.1, ** p<0.05, *** p<0.01

[1] Available readings of pollen count for the current day and the 6 previous days is used to calculate the weekly average of pollen. Then the result is divided by 100.

can be that the negative impact of air pollution/allergens on respiratory health is stronger on days which are on average warmer and also on those days the minimum temperature of which is lower.

Month indicators, similar to the previous results, show very significant effects on the number of out-patient hospital admissions among all age groups. All months from April to October estimate fewer number of admissions compared to March (baseline), ceteris paribus. Also, most of the weekdays are expected to have less total out-patient admissions when compared with Sunday which is our baseline. As explained earlier, the reason behind it can be that people select Sundays to go to hospitals since they are working on the other days. Again, the national holiday indicator shows no relationship with the number of admissions.

Table 3.6 shows the estimation results from running our model for only in-patient respiratory admissions, and the sub-groups of "under 5", "between 5 and 65", and "over 65", using the weekly average of pollen. Unlike the previous two models (total admissions and out-patient admissions only), surprisingly, in-patient admissions do not seem to have a significant dependence on the same-day level of $PM_{2.5}$. The first and second lag of $PM_{2.5}$, show some relationships with the number of in-patient admissions within the subgroups of "all patients" and "children under 5", but in general it seems that $PM_{2.5}$ is not causing respiratory problems that would need hospital admissions of more than 24 hours, unless people are exposed to it for a longer period of time. A similar result is derived after running the model with other specifications of pollen count[13]. This result is in contradiction to what Moeltner et al. (2013) find in their paper. The reason behind it can be the fact that in this paper we are using a longer time window.

Again, pollen count is not showing any significant effect on the number of admissions among any group or sub-group, which is similar to the results from running the model, taking the other 3 approaches towards pollen count[13].

Most of the air quality and meteorological factors are not showing strong association with in-patient admission counts within our four group(s)/sub-group(s). However, although not very significant, estimated coefficients of CO show a positive relationship between the CO level and the number of in-patient admissions in most of the groups. This can be considered as a support to the hypothesis that a bit of increase in the amount of CO might cause severe health issues that lead to in-patient admissions, rather than out-patient ones.

[13]Please see tables A.3, B.3, and C.3 in Appendix A, Appendix B, and Appendix C, respectively.

Table 3.6: Results for in-patient admissions - weekly average of pollen is used[1]

Variable	**All Patients**	**Under 5**	**Between 5 and 65**	**Over 65**
$PM_{2.5}$	-0.001	-0.006	-0.005	0.004
$PM_{2.5}$ (first lag)	0.005**	0.016**	0.006	0.002
$PM_{2.5}$ (second lag)	-0.005**	-0.022**	-0.005	-0.004
Pollen (weekly average[1]/100)	-0.001	0.010	0.003	-0.006
CO	0.099	1.217*	0.595*	-0.522*
NO2	0.004	0.025*	0.004	0.000
O3	1.295	-0.905	6.471**	-3.017
Avg. temperature	-0.014**	-0.021	-0.020**	-0.007
Dew point	0.005	0.005	0.016**	-0.005
Max. temperature	0.004	0.002	0.005	0.004
Min. temperature	0.004	0.010	0.002	0.004
Relative humidity	-0.004	-0.009	-0.011**	0.002
Avg. wind speed	-0.002	0.032*	-0.009	-0.002
Max. wind speed	0.004	-0.005	0.005	0.005
Precipitation	0.035	-0.178	0.206	-0.106
Avg. air pressure	-0.009***	-0.021**	-0.014***	-0.002
April	-0.223***	-0.278***	-0.233***	-0.194***
May	-0.261***	-0.623***	-0.239***	-0.200***
June	-0.319***	-0.907***	-0.309***	-0.223***
July	-0.495***	-1.305***	-0.439***	-0.419***
August	-0.600***	-1.273***	-0.551***	-0.524***
September	-0.431***	-1.064***	-0.275***	-0.452***
October	-0.347***	-1.143***	-0.174***	-0.355***
Monday	0.224***	0.338***	0.219***	0.212***
Tuesday	0.164***	0.140	0.230***	0.111**
Wednesday	0.023	0.041	0.039	0.009
Thursday	0.095***	0.028	0.088	0.112**
Friday	0.079**	0.014	0.067	0.100*
Saturday	-0.008	-0.096	0.043	-0.037
National holiday	-0.139	-0.352	-0.144	-0.111
Constant	10.079***	18.227**	13.761***	3.572
Number of Observations	1507	1507	1507	1507

* p<0.1, ** p<0.05, *** p<0.01

[1] Available readings of pollen count for the current day and the 6 previous days is used to calculate the weekly average of pollen. Then the result is divided by 100.

Month indicators, like before, show very significant effects on the number of total in-patient admissions among all age groups. All months from April to October estimate on average fewer number of in-patient admissions compared to March (baseline). Most of the day-of-week indicators, however, do not show significant effects on the dependent variable. This can mean that when the problem is too serious that need in-patient hospitalization, it does not matter what day of the week it is and people go to hospitals and get hospitalized, regardless. National holiday indicator shows no relationship with the number of in-patient admissions.

3.4.3 Marginal Effects of $PM_{2.5}$ on Treatment Costs

Health problems cause direct and indirect costs to societies. When an individual gets sick, the first adverse effect to society will be the loss of labor productivity for the period of time that he/she is being treated. The other costs to society are all of the direct and indirect costs of treating the patient. There are also costs associated with foregone recreational opportunities and other changes in daily activities in reaction to bad air (e.g. child cannot walk home form school and needs to be picked up, etc.)

According to the hospital admission data that we received from Nevada Center for Health Statistics and Informatics at the University of Nevada - Reno, the following is the average treatment cost of an out-patient admission (in 2015 prices) along with its standard deviation for the three age groups of "under 5", "between 5 and 65", and "above 65", respectively: \$1869.45 (SD: \$1781.94), \$2692.87 (SD: \$3614.78), and \$7242.62 (SD: \$7864.39). Using our results reported in table 3.5, for each 1 $\mu g/m^3$ increase in the level of $PM_{2.5}$, on average there will be about \$411 and \$169 increase in the daily total treatment cost of patients between 5 and 65 and those above 65 years old, respectively. Considering all patients, there will be an increase of around \$644 in the total daily treatment cost of respiratory out-patients in the Reno/Sparks area in Northern Nevada. Using these daily estimates, we can expect to witness about \$150,000, \$62,000, and \$235,000 increase in the annual total treatment cost of patients between 5 and 65, those above 65 years old, and all patients, respectively, due to 1 $\mu g/m^3$ increase in the level of $PM_{2.5}$.

3.5 Conclusion

Air pollution and allergens are known to be the reason for many respiratory illnesses. Existence of high concentrations of chemicals such as Carbon Monoxide (CO), Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), and/or Ozone (O3) in air on one hand, and presence of Atmospheric Particulate Matter (PM) of different diameters (especially $PM_{2.5}$) beyond specific thresholds, on the other hand, have been shown to cause a variety of health problems. However, there is not much research available on the health impact of $PM_{2.5}$ in existence of pollen to see whether or not the estimated effect of $PM_{2.5}$ is over-/under-estimated when the impact of pollen is neglected. The few studies which have addressed a similar research question, have mostly focused on asthma patients, especially among children and have studied the effect of PM_{10}, rather than $PM_{2.5}$[14]. Our main research question in this paper was to examine how important it is to control for pollen counts to obtain unbiased marginal effects of air pollutants (especially $PM_{2.5}$) on respiratory hospital admission counts/rates.

We used a Quasi-Maximum Likelihood (QMLE) approach which generates fully robust standard errors for the coefficients in the parameterized mean function to study the effects of $PM_{2.5}$ and pollen on the number of in-patient, out-patient, and total hospital admissions for respiratory health issues (Disease Codes (IDCs) 460.0-486.99, and 488.0-519.99, except for influenza (IDC 487.00-487.99)) among all age groups, under 5 years old, between 5 and 65, and over 65 years old. We focused our study on Reno/Sparks area in Northern Nevada, and used daily time-series data from 2009 to 2015 (from March to October of each year due to no pollen data collection from November to the end of February of the next year). In order to carry out our study in a way to produce reliable estimates, we controlled for a comprehensive group of air pollution and meteorological factors such as Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Ozone (O3), average daily temperature, min./max. daily temperature, average daily dew point, average wind speed, maximum wind speed, relative humidity, prescription, and average daily air pressure. We also controlled for the temporal effects: months of the year, days of the week, and also national holidays.

Our results showed a weak but significant positive relationship between the level of $PM_{2.5}$ and the number of total admissions and out-patient admissions within all of the age groups except children under 5 years old[15]. However, we found no sign of a relationship between $PM_{2.5}$ and

[14]The papers which are studying the above mentioned questions in the literature are reviewed in the section 3.2 of this chapter.

[15]One reason for not finding any significant relationship between $PM_{2.5}$ and the number of total admissions

the number of in-patient admissions within any group/subgroup. Also, estimated coefficients for the lags of $PM_{2.5}$ did not show much of association with the number of admissions in most of the age groups. Somewhat surprisingly, we did not find any evidence that shows any relationship between the pollen count and the number of total admissions or in-/out-patient admissions among any group/sub-group. In order to check for the robustness of our results derived from running our model using weekly average of pollen count, we ran our model using three other specifications for pollen as well: 1) daily pollen count with linear interpolation; 2) daily pollen count with step function; and 3) only those days that have pollen count readings are used for the analysis[16]. Their results also showed no relationship between the pollen count and the number of total admissions or in-/out-patient admissions among any group/sub-group.

As explained earlier, the pollen data we received from the only counting center in Reno did not provide us with a daily measurement of pollen. Instead, the data was collected on average once every 7 to 10 days, which is the most important shortcoming of our pollen data and can be the reason for not finding any relationship between the pollen count and admission counts, regardless of the approach taken to control for pollen effect. However, the main finding of our paper is that, acknowledging the shortcomings of the data, including all other control variables (weather, time effects, etc.) appears to be sufficient to obtain meaningful pollution effects without directly controlling for pollen for our application.

and out-patient admissions within children under 5 can be the smaller sample size for this age-group compared to the other groups.

[16]The results of running the model using these three approaches towards pollen are reported in Appendices A, B, and C, respectively.

Bibliography

Abramitzky, R., L. P. Boustan, and K. Eriksson (2012). A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. Technical report, National Bureau of Economic Research.

Anderson, H., A. P. de Leon, J. Bland, J. Bower, J. Emberlin, and D. Strachan (1998). Air pollution, pollens, and daily admissions for asthma in london 1987–92. *Thorax 53*(10), 842–848.

Banerjee, R. (2009). Income growth of new immigrants in canada: Evidence from the survey of labour and income dynamics. *Relations Industrielles/Industrial Relations*, 466–488.

Beenstock, M., B. R. Chiswick, and A. Paltiel (2005). Endogenous assimilation and immigrant adjustment in longitudinal data.

Beenstock, M., B. R. Chiswick, and A. Paltiel (2010). Testing the immigrant assimilation hypothesis with longitudinal data. *Review of Economics of the Household 8*(1), 7–27.

Borjas, G. (1995a). Assimilation and changes in cohort quality revisited: what happened to immigrant earnings in the 1980s? *Journal of labor economics 13*(2), 201.

Borjas, G. J. (1985). Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of labor Economics*, 463–489.

Borjas, G. J. (1989). Immigrant and emigrant earnings: A longitudinal study. *Economic inquiry 27*(1), 21–37.

Borjas, G. J. (1994). The economics of immigration. *Journal of economic literature 32*(4), 1667–1717.

Borjas, G. J. (1995b). The economic benefits from immigration. *The Journal of Economic Perspectives 9*(2), 3–22.

Borjas, G. J. (1999). The economic analysis of immigration. *Handbook of labor economics 3*, 1697–1760.

Borjas, G. J. (2013). The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again. Technical report, National Bureau of Economic Research.

Butry, D. T., E. Mercer, J. P. Prestemon, J. M. Pye, and T. P. Holmes (2001). What is the price of catastrophic wildfire? *Journal of Forestry 99*(11), 9–17.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64.

Card, D. (2005). Is the new immigration really so bad? *Economic Journal*, F300–F323.

Card, D. (2009). Immigration and inequality. Technical report, National Bureau of Economic Research.

Chiswick, B. R. (1978). The effect of americanization on the earnings of foreign-born men. *The journal of political economy*, 897–921.

Chiswick, B. R. (1980). *An analysis of the economic progress and impact of immigrants.* University of Illinois at Chicago Circle, Department of Economics.

Chiswick, B. R., Y. L. Lee, and P. W. Miller (2005a). Immigrant earnings: a longitudinal analysis. *Review of Income and Wealth 51*(4), 485–503.

Chiswick, B. R., Y. L. Lee, and P. W. Miller (2005b). A longitudinal analysis of immigrant occupational mobility: A test of the immigrant assimilation hypothesis1. *International Migration Review 39*(2), 332–353.

Chiswick, B. R. and W. Miller (2011). The negative assimilation of immigrants: A special case. *Industrial & Labor Relations Review 64*(3), 502–525.

De Mendonça, M. J. C., A. Sachsida, and P. R. Loureiro (2006). Estimation of damage to human health due to forest burning in the amazon. *Journal of Population Economics 19*(3), 593–610.

De Sario, M., K. Katsouyanni, and P. Michelozzi (2013). Climate change, extreme weather events, air pollution and respiratory health in europe. *European Respiratory Journal 42*(3), 826–843.

Delfino, R. J., S. Brummel, J. Wu, H. Stern, B. Ostro, M. Lipsett, A. Winer, D. H. Street, L. Zhang, T. Tjoa, et al. (2008). The relationship of respiratory and cardiovascular hospital admissions to the southern california wildfires of 2003. *Occupational and environmental medicine*.

Dennekamp, M. and M. J. Abramson (2011). The effects of bushfire smoke on respiratory health. *Respirology 16*(2), 198–209.

Dockery, D. W. and C. A. Pope (1994). Acute respiratory effects of particulate air pollution. *Annual review of public health 15*(1), 107–132.

Duleep, H. O. and D. J. Dowhan (2002). Insights from longitudinal data on the earnings growth of us foreign-born men. *Demography 39*(3), 485–506.

Erbas, B., J.-H. Chang, S. Dharmage, E. Ong, R. Hyndman, E. Newbigin, and M. Abramson (2007). Do levels of airborne grass pollen influence asthma hospital admissions? *Clinical & Experimental Allergy 37*(11), 1641–1647.

Fan, J., S. Li, C. Fan, Z. Bai, and K. Yang (2016). The impact of pm2. 5 on asthma emergency department visits: a systematic review and meta-analysis. *Environmental Science and Pollution Research 23*(1), 843–850.

Friedberg, R. M. (2001). The impact of mass migration on the israeli labor market. *Quarterly Journal of Economics*, 1373–1408.

Friedberg, R. M. and J. Hunt (1995). The impact of immigrants on host country wages, employment and growth. *The Journal of Economic Perspectives 9*(2), 23–44.

Gagliardi, L. and S. Lemos (2015). Evidence on immigrants assimilation into recipient labour markets using uk longitudinal data between 1981 and 2006. *Journal of Economic Geography*, lbv009.

Gleason, J. A., L. Bielory, and J. A. Fagliano (2014). Associations between ozone, pm 2.5, and four pollen types on emergency department pediatric asthma events during the warm season in new jersey: a case-crossover study. *Environmental research 132*, 421–429.

Hall, M., E. Greenman, and G. Farkas (2010). Legal status and wage disparities for mexican immigrants. *Social Forces*, 491–513.

Hanigan, I. and F. H. Johnston (2007). Respiratory hospital admissions were associated with ambient airborne pollen in darwin, australia, 2004–2005. *Clinical & Experimental Allergy 37*(10), 1556–1565.

Hu, W.-Y. (2000). Immigrant earnings assimilation: estimates from longitudinal data. *The American Economic Review 90*(2), 368–372.

Hunt, J. (2011). Which immigrants are most innovative and entrepreneurial? distinctions by entry visa. *Journal of Labor Economics 29*(3), 417–457.

Izquierdo, M., A. Lacuesta, and R. Vegas (2009). Assimilation of immigrants in spain: A longitudinal analysis. *Labour Economics 16*(6), 669–678.

Jalaludin, B., M. Smith, B. O'Toole, and S. Leeder (2000). Acute effects of bushfires on peak expiratory flow rates in children with wheeze: a time series analysis. *Australian and New Zealand journal of public health 24*(2), 174–177.

Jalaludin, B. B., B. I. OToole, and S. R. Leeder (2004). Acute effects of urban ambient air pollution on respiratory symptoms, asthma medication use, and doctor visits for asthma in a cohort of australian children. *Environmental Research 95*(1), 32–42.

Jayachandran, S. (2009). Air quality and early-life mortality evidence from indonesias wildfires. *Journal of Human resources 44*(4), 916–954.

Kaushal, N., Y. Lu, N. Denier, J. S.-H. Wang, and S. J. Trejo (2015). Immigrant employment and earnings growth in canada and the us: Evidence from longitudinal data. Technical report, National Bureau of Economic Research.

Kochi, I., P. A. Champ, J. B. Loomis, and G. H. Donovan (2012). Valuing mortality impacts of smoke exposure from major southern california wildfires. *Journal of forest economics 18*(1), 61–75.

LaLonde, R. J. and R. H. Topel (1992). The assimilation of immigrants in the us labor market. In *Immigration and the workforce: Economic consequences for the United States and source areas*, pp. 67–92. University of Chicago Press.

Lemos, S. (2013). Immigrant economic assimilation: Evidence from uk longitudinal data between 1978 and 2006. *Labour Economics 24*, 339–353.

Liu, J. C., G. Pereira, S. A. Uhl, M. A. Bravo, and M. L. Bell (2015). A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. *Environmental research 136*, 120–132.

Lofstrom, M. and J. J. Hayes (2011). H-1bs: How do they stack up to us born workers?

Long, J. E. (1980). The effect of americanization on earnings: some evidence for women. *Journal of Political Economy 88*(3), 620–629.

Lubotsky, D. (2007). Chutes or ladders? a longitudinal analysis of immigrant earnings. *Journal of Political Economy 115*(5), 820–867.

Mincer, J. A. (1974). Age and experience profiles of earnings. In *Schooling, experience, and earnings*, pp. 64–82. NBER.

Mithas, S. and H. C. Lucas Jr (2010). Are foreign it workers cheaper? us visa policies and compensation of information technology professionals. *Management Science 56*(5), 745–765.

Moeltner, K., M.-K. Kim, E. Zhu, and W. Yang (2013). Wildfire smoke and health impacts: A closer look at fire attributes and their marginal effects. *Journal of Environmental Economics and Management 66*(3), 476–496.

Nakamura, A. and M. Nakamura (1985). The second paycheck: An analysis of the employment and earnings of married women.

Picot, G. and P. Piraino (2013). Immigrant earnings growth: selection bias or real progress? *Canadian Journal of Economics/Revue canadienne d'économique 46*(4), 1510–1536.

Rittmaster, R., W. Adamowicz, B. Amiro, and R. Pelletier (2006). Economic analysis of health effects from forest fires. *Canadian Journal of Forest Research 36*(4), 868–877.

Rodríguez-Planas, N. and R. Vegas (2014). Do moroccan migrants to spain fare better or worse than other migrants? *Middle East Development Journal 6*(2), 308–328.

Shaposhnikov, D., B. Revich, T. Bellander, G. B. Bedada, M. Bottai, T. Kharkova, E. Kvasha, E. Lezina, T. Lind, E. Semutnikova, et al. (2014). Mortality related to air pollution with the moscow heat wave and wildfire of 2010. *Epidemiology 25*(3), 359–364.

Subiza, J., M. Jerezb, J. A. Jiméneza, M. J. Narganes, M. Cabrera, S. Varela, and E. Subiza (1995). Allergenic pollen and pollinosis in madrid. *Journal of Allergy and Clinical Immunology 96*(1), 15–23.

Tobias, A., I. Galan, and J. Banegas (2004). Non-linear short-term effects of airborne pollen levels with allergenic capacity on asthma emergency room admissions in madrid, spain. *Clinical & Experimental Allergy 34*(6), 871–878.

Tobias, A., I. Galan, J. Banegas, and E. Aranguez (2003). Short term effects of airborne pollen concentrations on asthma epidemic. *Thorax 58*(8), 708–710.

Vedal, S. and S. J. Dutton (2006). Wildfire air pollution and daily mortality in a large urban area. *Environmental research 102*(1), 29–35.

Youssouf, H., C. Liousse, L. Roblou, E.-M. Assamoi, R. O. Salonen, C. Maesano, S. Banerjee, and I. Annesi-Maesano (2014). Non-accidental health impacts of wildfire smoke. *International journal of environmental research and public health 11*(11), 11772–11804.

Appendices

Appendix A

Tables A.1, A.2, and A.3 present the results of running the model using daily pollen count for total admissions, out-patient admissions, and in-patient admissions, respectively. Linear interpolation technique is employed to fill the days with no pollen reading.

Table A.1: Results for the total admissions (both in-patient and out-patient) - daily pollen count is used[1]

Variable	**All Patients**	**Under 5**	**Between 5 and 65**	**Over 65**
$PM_{2.5}$	0.005***	0.001	0.005***	0.006**
$PM_{2.5}$ (first lag)	0.002*	-0.001	0.003**	0.005*
$PM_{2.5}$ (second lag)	-0.001	0.001	-0.000	-0.005*
Pollen[1]/100	-0.004	0.004	-0.004	-0.014*
CO	-0.293**	-0.148	-0.242*	-0.720***
NO2	0.006**	0.007	0.006**	-0.001
O3	-0.945	-0.411	-1.149	-1.310
Avg. temperature	0.005	0.005	0.008**	-0.001
Dew point	-0.001	0.003	-0.002	-0.005
Max. temperature	-0.003	-0.001	-0.005**	0.002
Min. temperature	-0.005**	-0.014***	-0.004	0.001
Relative humidity	-0.001	-0.004	0.000	0.002
Avg. wind speed	0.003	0.014**	0.000	-0.003
Max. wind speed	0.002	0.001	0.002	0.001
Precipitation	0.043	-0.088	0.089	-0.099
Avg. air pressure	-0.004*	-0.003	-0.004*	-0.003
April	-0.237***	-0.301***	-0.224***	-0.181***
May	-0.208***	-0.313***	-0.172***	-0.190***
June	-0.378***	-0.502***	-0.362***	-0.271***
July	-0.657***	-0.910***	-0.633***	-0.450***
August	-0.598***	-0.783***	-0.545***	-0.561***
September	-0.320***	-0.443***	-0.243***	-0.476***
October	-0.261***	-0.302***	-0.214***	-0.355***
Monday	0.049**	-0.095**	0.086***	0.125***
Tuesday	-0.001	-0.195***	0.055**	0.042
Wednesday	-0.064***	-0.213***	-0.018	-0.043
Thursday	-0.089***	-0.235***	-0.066***	0.032
Friday	-0.111***	-0.204***	-0.116***	0.046
Saturday	-0.093***	-0.144***	-0.080***	-0.063
National holiday	0.014	-0.033	0.045	-0.080
Constant	7.499***	5.567*	7.444***	4.607
Number of Observations	1660	1654	1654	1654

* p<0.1, ** p<0.05, *** p<0.01

[1] Linear interpolation technique is employed to fill the days with no reading.

Table A.2: Results for out-patient admissions - daily pollen count is used[1]

Variable	**All Patients**	**Under 5**	**Between 5 and 65**	**Over 65**
$PM_{2.5}$	0.006***	0.002	0.007***	0.011**
$PM_{2.5}$ (first lag)	0.002	-0.002	0.002*	0.008*
$PM_{2.5}$ (second lag)	0.000	0.002	0.000	-0.005
Pollen[1]/100	-0.005	0.003	-0.006	-0.033***
CO	-0.387***	-0.249	-0.359**	-1.196***
NO2	0.006**	0.007	0.006**	-0.000
O3	-1.260	-0.124	-2.043	2.871
Avg. temperature	0.010***	0.006	0.012***	0.010
Dew point	-0.003	0.003	-0.004	-0.009
Max. temperature	-0.005**	-0.001	-0.006**	-0.005
Min. temperature	-0.007***	-0.016***	-0.005*	0.001
Relative humidity	0.000	-0.004	0.002	0.005
Avg. wind speed	0.004	0.014*	0.001	-0.003
Max. wind speed	0.001	0.001	0.002	-0.007
Precipitation	0.040	-0.065	0.078	-0.055
Avg. air pressure	-0.003	-0.001	-0.003	-0.001
April	-0.244***	-0.307***	-0.226***	-0.158**
May	-0.201***	-0.287***	-0.167***	-0.189**
June	-0.398***	-0.467***	-0.373***	-0.377***
July	-0.705***	-0.875***	-0.666***	-0.532***
August	-0.605***	-0.750***	-0.548***	-0.643***
September	-0.301***	-0.400***	-0.240***	-0.558***
October	-0.238***	-0.249***	-0.223***	-0.336***
Monday	0.009	-0.132***	0.064**	0.002
Tuesday	-0.041*	-0.221***	0.029	-0.071
Wednesday	-0.084***	-0.231***	-0.024	-0.127*
Thursday	-0.131***	-0.252***	-0.088***	-0.086
Friday	-0.155***	-0.219***	-0.143***	-0.025
Saturday	-0.109***	-0.152***	-0.095***	-0.085
National holiday	0.056	-0.003	0.075	0.045
Constant	6.151***	4.078	6.150***	2.449
Number of Observations	1660	1660	1660	1660

* p<0.1, ** p<0.05, *** p<0.01

[1] Linear interpolation technique is employed to fill the days with no reading.

Table A.3: Results for in-patient admissions - daily pollen count is used[1]

Variable	**All Patients**	**Under 5**	**Between 5 and 65**	**Over 65**
$PM_{2.5}$	-0.001	-0.006	-0.005	0.003
$PM_{2.5}$ (first lag)	0.005**	0.016*	0.006*	0.003
$PM_{2.5}$ (second lag)	-0.006***	-0.025**	-0.005	-0.005
Pollen[1]/100	0.000	0.009	0.003	-0.004
CO	0.095	1.144*	0.502*	-0.443
NO2	0.003	0.019	0.003	-0.001
O3	0.177	-1.021	4.844*	-3.794
Avg. temperature	-0.015***	-0.019	-0.022***	-0.008
Dew point	0.006	0.004	0.017***	-0.002
Max. temperature	0.005*	0.002	0.005	0.006
Min. temperature	0.002	0.010	0.004	0.000
Relative humidity	-0.006**	-0.008	-0.012***	0.000
Avg. wind speed	-0.001	0.022	-0.004	-0.002
Max. wind speed	0.004	-0.001	0.003	0.005
Precipitation	0.024	-0.267	0.211	-0.120
Avg. air pressure	-0.009***	-0.022**	-0.013***	-0.003
April	-0.206***	-0.237**	-0.205***	-0.194***
May	-0.240***	-0.616***	-0.202***	-0.191***
June	-0.306***	-0.939***	-0.286***	-0.212***
July	-0.479***	-1.321***	-0.417***	-0.403***
August	-0.571***	-1.212***	-0.509***	-0.510***
September	-0.410***	-1.005***	-0.259***	-0.431***
October	-0.349***	-1.058***	-0.178***	-0.366***
Monday	0.225***	0.363***	0.230***	0.199***
Tuesday	0.167***	0.166	0.234***	0.109**
Wednesday	0.026	0.072	0.038	0.011
Thursday	0.092***	0.019	0.093*	0.103**
Friday	0.083**	0.061	0.077	0.092*
Saturday	-0.013	-0.085	0.034	-0.041
National holiday	-0.174*	-0.435	-0.160	-0.156
Constant	10.411***	19.017**	13.096***	4.721
Number of Observations	1654	1654	1654	1654

* p<0.1, ** p<0.05, *** p<0.01

[1] Linear interpolation technique is employed to fill the days with no reading.

Appendix B

Tables B.1, B.2, and B.3 present the results of running the model using daily pollen count for total admissions, out-patient admissions, and in-patient admissions, respectively. Step function technique is employed to fill the days with no pollen reading.

Table B.1: Results for the total admissions (both in-patient and out-patient) - daily pollen count is used[1]

Variable	All Patients	Under 5	Between 5 and 65	Over 65
$PM_{2.5}$	0.005***	0.001	0.005***	0.006**
$PM_{2.5}$ (first lag)	0.002*	-0.001	0.003**	0.005*
$PM_{2.5}$ (second lag)	-0.001	0.001	-0.000	-0.005*
Pollen[1]/100	-0.003	-0.003	-0.002	-0.009
CO	-0.297**	-0.152	-0.245*	-0.734***
NO2	0.006**	0.007	0.006**	-0.001
O3	-0.943	-0.330	-1.172	-1.344
Avg. temperature	0.005	0.005	0.008**	-0.001
Dew point	-0.001	0.003	-0.002	-0.005
Max. temperature	-0.003	-0.001	-0.005**	0.001
Min. temperature	-0.005**	-0.014***	-0.004	0.001
Relative humidity	-0.001	-0.004	0.000	0.002
Avg. wind speed	0.003	0.014**	0.000	-0.003
Max. wind speed	0.002	0.001	0.002	0.001
Precipitation	0.043	-0.092	0.091	-0.096
Avg. air pressure	-0.004**	-0.003	-0.005*	-0.003
April	-0.237***	-0.296***	-0.226***	-0.184***
May	-0.206***	-0.315***	-0.170***	-0.186***
June	-0.376***	-0.517***	-0.356***	-0.260***
July	-0.653***	-0.930***	-0.624***	-0.432***
August	-0.594***	-0.805***	-0.535***	-0.541***
September	-0.317***	-0.461***	-0.235***	-0.460***
October	-0.258***	-0.315***	-0.208***	-0.342***
Monday	0.049**	-0.094**	0.085***	0.125***
Tuesday	-0.001	-0.194***	0.055**	0.042
Wednesday	-0.064***	-0.212***	-0.018	-0.042
Thursday	-0.088***	-0.234***	-0.066***	0.033
Friday	-0.111***	-0.203***	-0.116***	0.047
Saturday	-0.093***	-0.144***	-0.079***	-0.062
National holiday	0.015	-0.037	0.047	-0.076
Constant	7.593***	5.534*	7.523***	4.896
Number of Observations	1660	1654	1654	1654

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

[1] Step function technique is employed to fill the days with no reading.

Table B.2: Results for out-patient admissions - daily pollen count is used[1]

Variable	All Patients	Under 5	Between 5 and 65	Over 65
$PM_{2.5}$	0.006***	0.002	0.007***	0.011**
$PM_{2.5}$ (first lag)	0.002	-0.002	0.002*	0.008*
$PM_{2.5}$ (second lag)	0.000	0.002	0.000	-0.005
Pollen[1]/100	-0.005	-0.004	-0.003	-0.022**
CO	-0.393***	-0.254	-0.363**	-1.226***
NO2	0.006**	0.007	0.006**	0.000
O3	-1.250	-0.036	-2.061	2.796
Avg. temperature	0.010***	0.007	0.012***	0.010
Dew point	-0.003	0.003	-0.004	-0.008
Max. temperature	-0.005**	-0.001	-0.006**	-0.006
Min. temperature	-0.007***	-0.016***	-0.005*	0.001
Relative humidity	0.000	-0.004	0.002	0.005
Avg. wind speed	0.004	0.014*	0.001	-0.003
Max. wind speed	0.001	0.001	0.002	-0.007
Precipitation	0.041	-0.069	0.079	-0.049
Avg. air pressure	-0.003	-0.001	-0.003	-0.002
April	-0.244***	-0.301***	-0.228***	-0.163**
May	-0.199***	-0.288***	-0.165***	-0.179**
June	-0.396***	-0.482***	-0.366***	-0.353***
July	-0.702***	-0.895***	-0.657***	-0.491***
August	-0.601***	-0.773***	-0.538***	-0.599***
September	-0.298***	-0.419***	-0.232***	-0.522***
October	-0.236***	-0.264***	-0.217***	-0.307***
Monday	0.009	-0.131***	0.063**	0.000
Tuesday	-0.040*	-0.219***	0.029	-0.069
Wednesday	-0.084***	-0.230***	-0.024	-0.125*
Thursday	-0.130***	-0.250***	-0.088***	-0.083
Friday	-0.155***	-0.219***	-0.143***	-0.024
Saturday	-0.109***	-0.151***	-0.095***	-0.083
National holiday	0.057	-0.008	0.077	0.052
Constant	6.276***	4.063	6.261***	3.145
Number of Observations	1660	1660	1660	1660

* p<0.1, ** p<0.05, *** p<0.01

[1] Step function technique is employed to fill the days with no reading.

Table B.3: Results for in-patient admissions - daily pollen count is used[1]

Variable	**All Patients**	**Under 5**	**Between 5 and 65**	**Over 65**
$PM_{2.5}$	-0.001	-0.006	-0.005	0.003
$PM_{2.5}$ (first lag)	0.005**	0.016*	0.006*	0.003
$PM_{2.5}$ (second lag)	-0.006***	-0.025**	-0.005	-0.005
Pollen[1]/100	0.002	0.007	0.007	-0.002
CO	0.098	1.161*	0.510*	-0.446
NO2	0.003	0.019	0.003	-0.001
O3	0.150	-0.997	4.789*	-3.808
Avg. temperature	-0.015***	-0.019	-0.022***	-0.008
Dew point	0.006	0.004	0.017***	-0.002
Max. temperature	0.005*	0.002	0.005	0.006
Min. temperature	0.002	0.010	0.003	0.000
Relative humidity	-0.006**	-0.008	-0.012***	0.000
Avg. wind speed	-0.001	0.022	-0.004	-0.002
Max. wind speed	0.004	-0.001	0.003	0.005
Precipitation	0.025	-0.270	0.213	-0.119
Avg. air pressure	-0.009***	-0.021**	-0.013***	-0.003
April	-0.208***	-0.237**	-0.208***	-0.195***
May	-0.240***	-0.619***	-0.203***	-0.189***
June	-0.302***	-0.947***	-0.279***	-0.209***
July	-0.474***	-1.333***	-0.410***	-0.397***
August	-0.565***	-1.225***	-0.500***	-0.503***
September	-0.405***	-1.016***	-0.252***	-0.426***
October	-0.345***	-1.066***	-0.173***	-0.361***
Monday	0.225***	0.364***	0.230***	0.198***
Tuesday	0.166***	0.166	0.232***	0.109**
Wednesday	0.026	0.071	0.037	0.011
Thursday	0.091***	0.017	0.092*	0.104**
Friday	0.083**	0.060	0.076	0.093*
Saturday	-0.013	-0.085	0.033	-0.041
National holiday	-0.173*	-0.438	-0.158	-0.154
Constant	10.376***	18.743**	12.970***	4.794
Number of Observations	1654	1654	1654	1654

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

[1] Step function technique is employed to fill the days with no reading.

Appendix C

Tables C.1, C.2, and C.3 present the results of running the model using daily pollen count for total admissions, out-patient admissions, and in-patient admissions, respectively. Only days with pollen count readings are used in the model for our analysis.

Table C.1: Results for the total admissions (both in-patient and out-patient) - daily pollen count is used[1]

Variable	All Patients	Under 5	Between 5 and 65	Over 65
$PM_{2.5}$	0.008***	0.006	0.009***	0.008
$PM_{2.5}$ (first lag)	0.005	0.004	0.005	0.008
$PM_{2.5}$ (second lag)	-0.003	-0.001	-0.003	-0.002
Pollen[1]/100	0.002	0.008	0.004	-0.013
CO	-0.469*	-0.140	-0.409	-1.120**
NO2	-0.002	-0.016	0.003	-0.007
O3	-3.169	-4.584	-3.500	0.131
Avg. temperature	0.004	-0.009	0.007	0.016
Dew point	-0.002	-0.003	0.003	-0.019**
Max. temperature	0.002	0.018***	-0.006	0.011
Min. temperature	-0.012**	-0.022**	-0.006	-0.019**
Relative humidity	-0.000	0.000	-0.004	0.015**
Avg. wind speed	0.002	0.001	0.002	-0.002
Max. wind speed	0.002	0.004	0.001	0.003
Precipitation	0.195**	-0.099	0.313***	0.020
Avg. air pressure	0.003	0.003	0.003	0.003
April	-0.346***	-0.420***	-0.336***	-0.262***
May	-0.277***	-0.431***	-0.250***	-0.150*
June	-0.463***	-0.503***	-0.469***	-0.347***
July	-0.703***	-0.954***	-0.721***	-0.359**
August	-0.629***	-0.848***	-0.585***	-0.509***
September	-0.376***	-0.516***	-0.311***	-0.446***
October	-0.246***	-0.293**	-0.244***	-0.154
Monday	0.161	0.086	0.171	0.255**
Tuesday	0.058	-0.110	0.081	0.218
Wednesday	0.058	0.021	0.029	0.266**
Thursday	0.016	0.046	-0.033	0.210
Friday	-0.430***	-0.802***	-0.463***	0.206
Constant	1.538	0.319	1.416	-1.350
Number of Observations	342	341	341	341

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

[1] Only days with pollen count readings are used in the model for our analysis.

Table C.2: Results for out-patient admissions - daily pollen count is used[1]

Variable	**All Patients**	**Under 5**	**Between 5 and 65**	**Over 65**
$PM_{2.5}$	0.009***	0.006	0.010***	-0.000
$PM_{2.5}$ (first lag)	0.007	0.001	0.007	0.030
$PM_{2.5}$ (second lag)	-0.002	0.000	-0.003	-0.005
Pollen[1]/100	-0.002	0.003	0.000	-0.052*
CO	-0.561*	-0.168	-0.627*	-1.091
NO2	-0.005	-0.021**	0.002	-0.015
O3	-3.733	-5.257	-4.533	10.376
Avg. temperature	0.006	-0.008	0.012	-0.014
Dew point	-0.002	-0.003	-0.002	-0.004
Max. temperature	0.002	0.022***	-0.008	0.035***
Min. temperature	-0.013**	-0.027***	-0.007	-0.022
Relative humidity	-0.000	0.001	-0.002	0.005
Avg. wind speed	0.002	0.003	0.002	-0.014
Max. wind speed	0.002	0.003	0.001	0.002
Precipitation	0.231**	-0.044	0.324**	0.124
Avg. air pressure	0.007	0.005	0.007	0.010
April	-0.368***	-0.426***	-0.358***	-0.193
May	-0.279***	-0.408***	-0.243***	-0.094
June	-0.491***	-0.478***	-0.493***	-0.438**
July	-0.765***	-0.913***	-0.750***	-0.479*
August	-0.671***	-0.824***	-0.615***	-0.675***
September	-0.379***	-0.473***	-0.320***	-0.620***
October	-0.246***	-0.232**	-0.249***	-0.234
Monday	0.143	0.082	0.159	0.260
Tuesday	0.036	-0.116	0.081	0.202
Wednesday	0.032	0.022	0.028	0.189
Thursday	-0.014	0.020	-0.052	0.303
Friday	-0.502***	-0.717***	-0.464**	0.045
Constant	-1.763	-1.792	-2.020	-8.867
Number of Observations	342	342	342	342

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

[1] Only days with pollen count readings are used in the model for our analysis.

Table C.3: Results for in-patient admissions - daily pollen count is used[1]

Variable	**All Patients**	**Under 5**	**Between 5 and 65**	**Over 65**
$PM_{2.5}$	0.007	0.011	0.000	0.012*
$PM_{2.5}$ (first lag)	-0.002	0.060	-0.007	-0.005
$PM_{2.5}$ (second lag)	-0.005	-0.045	-0.008	-0.000
Pollen[1]/100	0.020*	0.058**	0.027*	0.006
CO	-0.136	0.317	0.850	-1.128*
NO2	0.008	0.053*	0.013	-0.003
O3	-0.586	7.692	3.775	-5.532
Avg. temperature	-0.002	-0.035	-0.032*	0.030*
Dew point	0.002	-0.000	0.035**	-0.026**
Max. temperature	0.004	-0.022	0.013	-0.000
Min. temperature	-0.008	0.039	-0.006	-0.018
Relative humidity	0.001	-0.010	-0.018*	0.019**
Avg. wind speed	0.005	0.010	0.008	0.005
Max. wind speed	0.002	-0.007	-0.001	0.004
Precipitation	0.043	-2.373	0.226	-0.030
Avg. air pressure	-0.011	-0.023	-0.021*	-0.001
April	-0.262***	-0.306	-0.207*	-0.293**
May	-0.269***	-0.578*	-0.292**	-0.171
June	-0.353***	-0.790*	-0.326**	-0.288*
July	-0.450***	-1.201**	-0.494**	-0.284
August	-0.465***	-1.107*	-0.395**	-0.412**
September	-0.376***	-1.032**	-0.269	-0.350**
October	-0.253***	-1.142***	-0.219	-0.124
Monday	0.234*	0.145	0.229**	0.243
Tuesday	0.163	0.104	0.104	0.226
Wednesday	0.163	0.027	0.018	0.302
Thursday	0.127	0.332	0.059	0.144
Friday	-0.132	-14.631***	-0.542***	0.295
Constant	12.070*	21.181	20.199*	1.525
Number of Observations	341	341	341	341

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

[1] Only days with pollen count readings are used in the model for our analysis.